

Les ateliers de l'éthique The Ethics Forum

Les ateliers de l'éthique
The Ethics Forum

Le réalisme moral

Peter Railton

Volume 11, Number 2-3, Fall 2016

URI: <https://id.erudit.org/iderudit/1041772ar>

DOI: <https://doi.org/10.7202/1041772ar>

[See table of contents](#)

Publisher(s)

Centre de recherche en éthique (CRÉ)

ISSN

1718-9977 (digital)

[Explore this journal](#)

Cite this article

Railton, P. (2016). Le réalisme moral. *Les ateliers de l'éthique / The Ethics Forum*, 11(2-3), 171–212. <https://doi.org/10.7202/1041772ar>

© Centre de recherche en éthique (CRÉ), 2017



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

LE RÉALISME MORAL

PETER RAILTON
UNIVERSITY OF MICHIGAN

TRADUIT PAR DENIS COURVILLE
CANDIDAT AU DOCTORAT EN PHILOSOPHIE, UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Il existe une tendance, parmi les philosophes contemporains, à considérer le scepticisme moral comme une position irrésistible, même chez ceux pour qui le scepticisme envers les sciences empiriques ne représente absolument pas une position convaincante. Pour diverses raisons, parmi lesquelles on trouve une méfiance compréhensible envers l'absolutisme moral, on considère comme un signe de bon sens d'expliquer [*explain away*] toute apparence d'objectivité dans le domaine du discours moral. Il est devenu si commun de traiter la moralité comme subjective ou conventionnelle dans le cadre d'une culture intellectuelle séculière que la plupart d'entre nous éprouve de la difficulté à s'imaginer ce à quoi pourrait ressembler l'existence de faits auxquels les jugements moraux sont redevables.

Loin d'être découragés, certains philosophes ont tenté de rendre compte de l'objectivité de la moralité en faisant valoir que la raison ou la science sert de fondement à l'éthique. Toutefois, l'histoire des tentatives de ce type n'inspire guère la confiance. Bien que le rationalisme en éthique ait retenu certains adhérents bien après que les autres rationalismes aient été abandonnés, les courants philosophiques puissants qui ont érodé l'idée d'une raison qui pourrait être, à elle seule, à même de nous offrir une position en fonction de laquelle des conclusions substantielles pourraient être dérivées ne manifestent aucun signe de fléchissement. Qui plus est, l'éthique naturaliste n'est pas parvenue jusqu'à présent à formuler une synthèse plausible des champs empirique et normatif : plus elle s'adonne à des analyses descriptives au sujet de l'origine des normes, moins elle retient sa force morale que l'on reconnaît; plus elle tente d'offrir une base reconnaissable à la critique ou à la reconstruction morale, moins elle retient un lien solide avec les théories psychologiques ou sociales descriptives.¹

Dans ce qui suit, je présenterai de manière programmatique une forme de naturalisme éthique qui, bien que considérablement redevable aux idées des théoriciens antérieurs, cherche à établir une relation plus satisfaisante entre les champs normatif et empirique. Cette relation, je crois, ne peut être démontrée par aucune preuve. Mon objectif ne consiste pas à réfuter le scepticisme moral, tout comme l'objectif des naturalistes épistémiques contemporains ne consiste pas à réfuter le scepticisme cartésien. Les prétentions du naturaliste, dans un cas comme dans l'autre, sont plus modestes. Il tente d'abord d'offrir une analyse de l'épistémologie ou de l'éthique permettant de saisir comment les fonctions évaluatives centrales dans ces domaines pourraient être accomplies dans le cadre des théories empiriques actuelles (ou prospectives). Il tente ensuite de montrer comment les explications non naturalistes traditionnelles s'appuient sur des suppositions qui sont, d'une certaine manière, incohérentes, ou qui s'accordent mal avec l'état actuel des sciences. Troisièmement, il lance un certain défi au sceptique, en montrant notamment comment une explication sceptique de nos pratiques épistémiques et morales peut être aussi plausible, utile ou intéressante que l'explication proposée par le naturalisme, et comment une reconstruction de telles pratiques – si le sceptique devait tenter d'en offrir une, comme il lui arrive souvent de le faire – peut parvenir à conserver la place et la fonction typiques de ces pratiques dans l'ordre des questions humaines. Je me concentrerai principalement sur la première de ces trois aspirations.

Une chose doit d'emblée être affirmée. Certains peuvent être interpellés ou repoussés par le réalisme moral sous prétexte qu'il s'agit là d'une conception de l'éthique qui est le plus à même d'exprimer le sérieux d'une haute considération pour la moralité. Or la moralité peut être prise au sérieux, même de manière excessive, sans qu'elle implique le réalisme moral. En effet, une objection possible pouvant être adressée au type de réalisme moral que je défends ici maintient que cette position peut ne pas considérer la moralité avec suffisamment de sérieux.

I. LES ESPÈCES DE RÉALISME MORAL

Diverses positions ont affirmé être – ou sont accusées d'être – réalistes en matière de moralité, de sorte qu'une caractérisation initiale de la position que je défends ici est requise avant que l'on puisse poursuivre davantage l'analyse. Les affirmations – et les accusations – de réalisme moral se fondent généralement sur certaines ou l'ensemble des questions suivantes. Grosso modo : 1) Le cognitivisme – Est-ce que tous les jugements moraux sont capables de vérité et de fausseté? 2) Les théories de la vérité – Si les jugements moraux ont des valeurs de vérité, en quel sens est-ce le cas? 3) L'objectivité – De quelles façons, s'il en est, l'existence de propriétés morales dépend-elle des états d'esprit actuels ou possibles d'êtres intelligents? 4) Le réductionnisme – Les propriétés morales sont-elles réductibles à – ou surviennent-elles en un sens plus faible – des propriétés non morales? 5) Le naturalisme – Les propriétés morales sont-elles des propriétés naturelles? 6) L'empirisme – En venons-nous à une connaissance des faits moraux de la même façon dont nous en venons à connaître les faits des sciences empiriques, ou sont-ils révélés par la raison ou par un certain mode d'appréhension spécifique? 7) La bivalence – est-ce que le principe du tiers exclu s'applique aux jugements moraux? 8) La déterminabilité – Quelle que soit la procédure dont nous disposons pour évaluer les jugements moraux, quelle proportion de la moralité peut vraisemblablement être déterminable? 9) La catégoricité – Est-ce que tous les agents rationnels ont nécessairement des raisons pour obéir à des impératifs moraux? 10) L'universalité – Les impératifs moraux s'appliquent-ils à tous les agents rationnels, même – s'il en existe – à celles et ceux qui ne disposent d'aucune raison pour s'y conformer? 11) L'évaluation des modalités existantes – Est-ce que les croyances morales actuelles sont approximativement vraies, ou est-ce que les intuitions morales qui prévalent constituent, en un tout autre sens, des données privilégiées? 12) Le relativisme – Est-ce que la vérité ou la justification des jugements moraux dépend directement des normes ou des pratiques individuellement – ou socialement – adoptées? 13) Le pluralisme – existe-t-il une seule forme de la vie bonne ou un seul code moral juste (*right moral code*), ou est-ce que des formes de vie ou des codes moraux différents pourraient être appropriés dans différentes circonstances?

Voici alors les coordonnées approximatives de ma propre position au sein de cet espace conceptuel à dimensions multiples. Je veux défendre une forme de réalisme qui fait valoir que des jugements moraux peuvent être des porteurs de valeurs de vérité [*can bear truth values*] en un sens fondamentalement non épis-

témique; que les propriétés morales sont objectives bien que relationnelles; que les propriétés morales surviennent sur les propriétés naturelles et peuvent se réduire à celles-ci; que l'analyse morale relève de la recherche empirique; que l'on ne peut pas savoir à priori si la bivalence s'applique aux jugements moraux ou la façon dont de tels jugements peuvent être évalués de façon déterminante; qu'il y a des raisons de penser que nous connaissons une part importante de la moralité, mais aussi qu'il y a des raisons de croire que les conceptions morales actuelles sont fausses à certains égards et pourraient même être fausses de façon très générale; qu'un agent rationnel pourrait n'avoir aucune raison d'obéir à des impératifs moraux, bien que ceux-ci puissent néanmoins s'appliquer à elle ou à lui; et que, bien que l'évaluation morale puisse reposer sur des critères parfaitement généraux, et en vertu de la nature même de ces critères, aucun type de vie n'est susceptible de convenir à tous les individus, de même qu'aucun ensemble de normes ne peut convenir à toutes les sociétés et à toutes les époques. La position ainsi décrite pourrait bien se nommer le « réalisme moral complètement délirant » [*stark, raving moral realism*], mais pour des raisons de syntaxe, je la nommerai fadement « réalisme moral ». L'utilisation de ces expressions n'est pas exclusive à la position décrite ici. D'autres positions, qui occupent des coordonnées plus ou moins distinctes, pourraient revendiquer l'une de ces deux appellations.

II. LA DISTINCTION FAIT-VALEUR

Toute tentative visant à justifier une conception naturaliste du réalisme moral fonce tête baissée vers la distinction fait-valeur. Les diverses formulations de cette distinction, tout comme les arguments justifiant sa raison d'être, ont déjà été présentés dans le passé, mais j'insisterai, en l'occurrence, sur quelques questions concernant le statut épistémique et ontologique des jugements de valeur par opposition aux jugements de fait.

Il semble que l'argument le plus souvent évoqué en faveur de la distinction fait-valeur est d'ordre épistémique : on affirme que les conflits sur des questions de valeur peuvent persister même lorsque tous les moyens d'adjudication rationnels et scientifiques ont été déployés, de sorte que les jugements de valeur ne peuvent être cognitifs au même titre que le sont les jugements factuels ou logiques. Cette thèse est, en partie, défendue en invoquant le caractère instrumental (hypothétique) de la raison, lequel empêcherait la raison de dicter les valeurs ultimes. En principe, cet argument affirme que deux individus qui diffèrent sur des valeurs ultimes pourraient, sans faire preuve de lacunes rationnelles, tenir ferme à leurs valeurs opposées, peu importe le nombre d'arguments ou de preuves mis de l'avant. À l'instar d'Ayer, « nous estimons que la discussion est possible dans les questions de morale, seulement si un système de valeurs est présupposé. »²

On pourrait tenter de contrer cette conclusion en remettant en cause la conception instrumentale de la rationalité. Mais malgré tous ses défauts et tout ce qu'elle se doit de développer, la conception instrumentale de la rationalité semble, à mon avis, la notion la plus claire dont on dispose concernant ce que

signifient pour un agent d'avoir des raisons d'agir. De plus, elle rend compte d'une caractéristique normative centrale du fait de donner des raisons [*reason-giving*], étant donné que nous pouvons aisément reconnaître la force impérieuse de l'idée qui s'impose à l'agent, selon laquelle un acte quelconque pourrait favoriser la réalisation de ses fins. Il serait difficile de faire sens de quelqu'un qui affirmerait sincèrement à la fois avoir certaines fins et, pourtant, qu'elles ne constituent pas pour lui des raisons *prima facie* suffisantes d'agir. (Bien sûr, il pourrait également croire qu'il a d'autres raisons suffisantes, voire compensatoires).

Pourtant, cette version de l'argument épistémique en faveur de la distinction fait-valeur est aux prises avec des difficultés même si l'on concède la conception instrumentale de la rationalité. Du point de vue de la raison instrumentale, la formation de croyances n'est qu'une activité parmi d'autres : dans la mesure où nous avons des raisons de nous y consacrer ou de l'entreprendre d'une manière plutôt qu'une autre, il s'agit au fond de questions concernant leur contribution à nos fins.³ Ce qui, pour un individu, est rationnel de croire sur la base d'une quelconque expérience variera non seulement en fonction de ses autres croyances, mais aussi en fonction de ce qu'il désire.⁴ Il s'ensuit qu'aucune argumentation ou expérience, à elle seule, ne peut contraindre quelqu'un sous peine d'irrationalité d'accepter même les énoncés factuels des sciences empiriques. Le débat entamé depuis déjà très longtemps sur la logique inductive illustre bien qu'un choix rationnel parmi plusieurs hypothèses divergentes requiert des critères plus riches et plus controversés en ce qui concerne le choix d'une théorie que ceux pouvant être seulement dégagés à partir de la raison instrumentale. Malheureusement pour l'opposition qu'Ayer désirait établir, nous considérons que l'argumentation sur des questions scientifiques n'est possible que si un système de valeurs est présupposé.

Hume a toutefois trouvé un moyen d'établir la distinction entre faits et valeurs sans recourir à l'idée suivant laquelle l'induction – ou même la déduction – pourrait exiger de la part d'un agent rationnel l'adoption de certaines croyances plutôt que d'autres lorsque cela entre en conflit avec ses fins contingentes.⁵ Car Hume met de l'avant la thèse selon laquelle la moralité est pratique, à savoir que si les faits moraux existent, ils offrent nécessairement pour tout être rationnel une raison (quoique peut-être non pas une raison décisive) d'agir moralement, quoi qu'il en soit des désirs particuliers propres à chacun. Si cette thèse est retenue à titre de prémisse, la conception instrumentale de la rationalité peut être probante après tout, puisqu'elle exclut la possibilité qu'il y ait des raisons catégoriques de ce type. En revanche, Hume ne présuppose pas que cela soit constitutif de la logique ou de la science que les faits mis en évidence par celles-ci s'imposent en tant que force catégorique à tout agent rationnel, de sorte que l'existence de faits logiques ou scientifiques, à la différence de faits moraux, est compatible avec le caractère instrumental de la raison.

Or cette façon de formuler la distinction fait-valeur n'est pas aussi convaincante que la thèse suivant laquelle la moralité est essentiellement pratique au sens où

l'entend Hume.⁶ Hume a sûrement raison lorsqu'il affirme qu'il y a une relation intrinsèque, sans aucun doute complexe, entre le fait d'accorder de la valeur [*valuing*] à quelque chose et le fait d'adopter une certaine forme d'attitude positive à son égard, laquelle fournit alors à quelqu'un une raison instrumentale d'agir. Nous ne croirions tout simplement pas quelqu'un qui affirmerait accorder une importance à l'honnêteté et qui, pourtant, ne montrerait jamais la moindre envie d'agir honnêtement lorsqu'il en aurait facilement l'occasion. Or il s'agit là d'un fait concernant la relation entre les valeurs *épousées* par un individu et ses raisons d'agir, et non pas d'un fait témoignant d'une relation entre l'évaluation morale et la motivation rationnelle.

Supposons, par exemple, que nous acceptions la caractérisation de Hume selon laquelle la justice est une vertu artificielle, dirigée vers le bien-être [*welfare*] général. Cela est, en un sens manifeste, une notion évaluative ou normative – « une valeur » au sens large tel qu'elle est employée dans le cadre de tels débats – et pourtant, il ne suit certainement pas de cette définition que tout être rationnel qui croit qu'un acte quelconque est juste en ce sens, et ce, peu importe ses désirs, aurait une raison instrumentale d'accomplir cet acte. Un individu rationnel peut ne pas accorder de valeur [*valuing*] à la justice pour elle-même [*for its own sake*], et peut avoir des fins contraires à celle-ci. Lorsque Hume discute de notre « obligation intéressée » à être juste, il semble reconnaître que, en fin de compte, il peut ne pas être possible de démontrer qu'un « fripon sensé » [*sensible knave*] a des raisons d'être juste. Bien entendu, Hume fait valoir que le reste d'entre nous – dont les cœurs se rebellent contre l'attitude du fripon rusé, lequel considère qu'il est acceptable de manquer à sa parole, de tricher ou de voler lorsque cela lui convient – a des raisons d'être juste, de considérer l'attitude du fripon comme injuste, et de se protéger contre ses prédatations.⁷

Le fripon pourrait affirmer, pour sa part, peut-être parce qu'il accepte l'analyse de la justice de Hume : « Oui, mon attitude est injuste ». Sur la base de la conception humienne de la relation entre la raison et les passions, le même fripon pourrait ajouter : « Mais qu'est-ce que cela peut me faire? », et ce, bien qu'il puisse être à même de saisir le contenu de son assertion précédente. Supposons que le fripon n'ait aucun doute quant à l'intelligibilité ou à la réalité du « bien-être général », et qu'il pense qu'il est parfaitement compréhensible que les individus attachent une grande importance à la notion de justice qui lui est corrélée dans le cadre de la vie en commun. Il réalise également que pour la majorité de l'humanité, dont les passions diffèrent des siennes, le fait d'être juste est une source et une condition de ce qui a le plus de valeur [*worthwhile*] dans la vie. Il comprend ainsi que les appels à la justice ont généralement une force motivante. En outre, il a lui-même recours à la catégorie de justice afin d'analyser le monde social, et il reconnaît – car ses calculs de fripon, en effet, en tiennent compte – la distinction entre, d'une part, les individus et les institutions qui sont véritablement justes et, de l'autre, ceux qui sont justes simplement en apparence ou sont communément considérés comme justes. Le fripon considère un bon nombre de concepts qui sont monnaie courante [*wide currency*] – les concepts religieux, par exemple – comme de simples fictions qui s'en prennent aux esprits

faibles, mais il ne considère pas la justice ainsi. Les esprits faibles et les moralistes, croit-il, ont couvert la justice de certains mythes – que la justice est sa propre récompense, que lorsque l'on voit ce qui est juste on aura automatiquement des raisons de faire ce qui est juste, et ainsi de suite. Mais cela étant dit, il croit que les esprits faibles et les moralistes ont également couvert la richesse et le pouvoir de certains mythes – que les riches ne sont pas véritablement heureux, que les puissants se précipitent vers leur chute, et ainsi de suite – et il ne doute pas, de ce point de vue, qu'il y ait des choses telles que la richesse et le pouvoir. Le fripon est heureux d'être affranchi des mythes prédominants concernant la richesse, le pouvoir et la justice; il est également heureux d'être libéré du fait de prêter autant ou peu d'attention, dans son esprit, à ces attributs, comme le commandent ses désirs ou ses circonstances. Il pourrait, par exemple, être convaincu du conseil de Mae West : il vaut tout à fait la peine d'acquérir des diamants, et « le bien [*goodness*] n'a rien à voir avec ça. »

Nous devons donc distinguer le travail consistant à déterminer ce à quoi un individu accorde de la valeur de celui visant à affirmer ce en quoi consiste pour un individu de prendre des mesures [*measurements*] à la lumière des critères d'une espèce d'évaluation qu'il reconnaît comme légitime.⁸

En niant la thèse de Hume concernant le caractère pratique du jugement moral et en supprimant par là même le fondement de la distinction humienne entre faits et valeurs, on ne nie pas pour autant le fait que la moralité possède cette caractéristique consistant à guider nos actions. La moralité peut certainement demeurer prescriptive au sein d'un cadre instrumental et peut se recommander à nous de la même façon dont, par exemple, l'épistémologie le peut : plusieurs fins humaines significatives et persistantes – bien que peut-être pas universelles – peuvent être poursuivies si nous appliquons certains critères d'évaluation à nos actions. Cela peut suffire à justifier à nous-mêmes notre souci persistant pour le statut épistémique ou moral de ce que nous faisons.⁹

En faisant valoir que la raison ne nous contraint pas à adopter des croyances ou des pratiques spécifiques, mises à part nos fins contingentes et variables, il peut sembler que je ne sois pas parvenu à composer avec le relativisme épistémique, et que j'aie miné l'argument en faveur du réalisme moral avant même de l'avoir présenté. La rationalité devient effectivement relative lorsqu'elle est instrumentale, mais cela n'est pas également le cas de l'épistémologie. La justification épistémique [*epistemic warrant*] de la croyance d'un individu peut être dissociée de la rationalité impliquée dans le fait d'entretenir cette croyance, car la justification épistémique peut être liée à un critère externe – comme c'est le cas, par exemple, dans le cadre des théories de la connaissance causales ou fiabilistes [*reliabilist*].¹⁰ Nous ne devons pas confondre l'idée selon laquelle nos fins en général sont bien servies par des mécanismes causaux et fiables de formation des croyances avec la thèse internaliste selon laquelle la raison nous contraint à adopter des moyens de ce type. Aux mécanismes fiables correspondent des coûts tout comme des bénéfices, et la poursuite efficace de certaines fins – le fripon ferait ici allusion à certaines fins religieuses, et à celles de certains moralistes – peut,

à certains égards, être incompatible avec l'adoption de moyens fiables d'investigation [*reliable means of inquiry*].

Cette réfutation de l'accusation de relativisme invite les partisans de la distinction fait-valeur à se déplacer sur un terrain ontologique. Peut-être que les faits et les valeurs ne peuvent représenter les pôles opposés d'une division épistémologique, marquée par ce que la raison et l'expérience peuvent nous interpeller à accepter. Néanmoins, l'idée de mécanismes causaux et fiables à la base de l'apprentissage moral et de faits moraux « dans le monde » sur lequel ces mécanismes agissent est sans doute si bizarre qu'il semble que je n'aie peut-être que multiplié mes difficultés.

III. LE RÉALISME DES VALEURS

Certes, l'idée d'une interaction causale avec la réalité morale serait intolérablement étrange si les faits moraux étaient considérés comme *sui genesis*.¹¹ Mais l'idée qu'il y a des mécanismes causaux à la base de l'apprentissage des faits moraux n'a rien de forcément bizarre si de tels faits sont constitués par des faits naturels. Telle est précisément la position que j'examinerai à présent. La réponse demeurera, toutefois, peu convaincante jusqu'à ce que l'on présente certains arguments positifs en faveur du réalisme à propos des faits moraux. À présent, tournons ainsi notre attention sur cette tâche.

Ce que l'on pourrait appeler « le stratagème générique du réalisme naturaliste » consiste à postuler un domaine de faits sur la base de la contribution qu'ils pourraient apporter à l'explication à posteriori de certains traits caractéristiques de notre expérience. Par exemple, un monde externe est postulé [*posited*] afin d'expliquer la cohérence, la stabilité et l'intersubjectivité de l'expérience sensible [*sense-experience*]. Un réaliste moral qui tire profit de cette stratégie doit montrer que la postulation de faits moraux peut, de manière similaire, avoir une fonction explicative. La stratégie peut réussir dans les deux cas seulement si la réalité postulée possède ces deux caractéristiques :

L'indépendance : La réalité existe et possède certains traits déterminés indépendamment du fait que nous pensions qu'elle existe ou qu'elle possède ces traits, voire indépendamment du fait que nous ayons de bonnes raisons de croire en cette réalité;

La rétroaction [*feedback*] : La réalité est constituée – et nous sommes constitués – de telle sorte que nous pouvons interagir avec elle, et ces interactions exercent de manière appropriée une influence et un contrôle constitutifs sur nos perceptions, nos pensées et nos actions.

Ces deux caractéristiques permettent au postulat [*posit*] réaliste de jouer un rôle dans l'explication de notre expérience, qui ne peut être remplacée sans perte par une simple *conception* de nous-mêmes ou de notre monde. Car bien que la médiation opérée par notre cadre conceptuel s'effectue même au niveau de nos

expériences perceptuelles les plus élémentaires, une réalité qui transcende l'expérience peut de multiples manières se faire sentir [*making itself felt*] causalement sans la permission de notre cadre conceptuel. Le succès ou l'échec de nos plans et de nos projets ne sont pas déterminés seulement par nos attentes. En résistant ou en cédant à des efforts mondains par des moyens qui ne sont pas anticipés par notre cadre conceptuel actuel, une réalité externe qui ne se révèle pourtant jamais directement dans la perception peut néanmoins influencer considérablement l'évolution subséquente de ce schème.

Le fait que le réalisme ait recours à un monde externe afin d'expliquer l'expérience sensorielle a souvent été critiqué sous prétexte qu'il ne s'agit là que d'une simple image. Mais avons-nous même une simple idée de ce à quoi peut ressembler une explication réaliste dans le cas des valeurs?¹² Je tenterai d'en esquisser une en présentant d'abord une conception de la valeur non morale : l'idée suivant laquelle quelque chose est désirable pour quelqu'un ou est bon [*good*] pour cet individu.¹³

Considérons d'abord l'idée d'*intérêt subjectif* propre à quelqu'un – ses désirs [*wants or desires*], conscients ou inconscients. Un intérêt subjectif peut être considéré comme une qualité secondaire, telle que le goût. Que je m'intéresse à quelque chose [*take a subjective interest in something*] revient à dire que cette chose a une *valeur* [*valence*] positive pour moi, c'est-à-dire que, dans des circonstances ordinaires, elle excite une attitude ou une inclination positive (pas nécessairement consciente) en moi. De la même manière, que j'affirme que le sucre me semble sucré revient à dire que, dans des circonstances normales, le sucre excite une certaine sensation gustative en moi. En tant que qualités secondaires, l'intérêt subjectif et la sucrosité éprouvée surviennent sur les qualités premières de celui qui perçoit, de l'objet (ou autre phénomène) perçu et du contexte environnant : celle ou celui qui perçoit est constitué de telle sorte qu'un tel objet dans un tel contexte excitera une sensation de ce type. Nous nommons la *base réductrice* de la qualité secondaire cet ensemble complexe de qualités premières, relationnelles et dispositionnelles.

Dans cette base réductrice s'offre à nous une notion objective qui correspond aux intérêts subjectifs et nous permet de les expliquer. Mais elle ne peut pas servir de fondement plausible à la notion de bien [*goodness*] non moral, étant donné que les intérêts subjectifs qu'elles fondent ne disposent pas d'une force normative suffisante pour rendre compte de l'idée de désirabilité [*desirableness*]. Les intérêts subjectifs sont souvent le reflet de l'ignorance, de la confusion ou du manque de considération, comme on peut le constater avec un certain recul. Le fait que je sois actuellement ainsi constitué que je désire quelque chose que, si j'en avais une meilleure connaissance, je souhaiterais n'avoir jamais désiré ne semble pas se présenter à moi comme ce qui fait partie de mon bien [*good*].

Pour remédier à ce défaut, permettons-nous d'introduire la notion d'*intérêt subjectif objectifié* d'un individu *A*, comme suit.¹⁴ Accordons à un individu *A* actuel des pouvoirs cognitifs et imaginatifs inconditionnels, de même qu'une

pleine connaissance de toute information factuelle et nomologique concernant sa constitution physique et psychologique, ses capacités, ses circonstances, son histoire, etc. *A* deviendra alors *A+*, lequel possède une connaissance complète et vive [*vivid*] de lui-même et de son environnement, et dispose d'une rationalité instrumentale qui ne peut en aucun cas faire défaut. Nous demandons ensuite à *A+* de nous dire non pas ce qu'*il* désire actuellement, mais ce qu'il voudrait que sa version non idéalisée de lui-même, *A*, veuille – ou, plus généralement, convoite – s'il devait se trouver lui-même dans la condition et les circonstances actuelles de *A*.¹⁵ Tout comme nous avons présupposé qu'il y a une base réductive au fondement des intérêts subjectifs actuels d'un individu *A*, nous présupposons qu'il y a une base réductive au fondement de ses intérêts subjectifs objectifiés, à savoir ces faits à propos de *A* et de ses circonstances que *A+* combinerait avec ses connaissances générales afin d'en arriver à des idées au sujet de ce qu'il voudrait vouloir s'il devait se trouver à la place de *A*.

Prenons, par exemple, Lonnie, un voyageur qui se trouve dans un pays étranger et qui se sent mal en point. Il désire avant tout venir à bout de son malaise et soulager son estomac, et il constate qu'il a un désir insatiable pour un goût familier : un grand verre de lait. Le lait est désiré par Lonnie, mais s'agit-il également de quelque chose de désirable pour lui? Lonnie-Plus peut voir que ce qui ne va pas avec Lonnie, outre le mal du pays, est la déshydratation, un mal qu'ont en commun plusieurs touristes, mais qui est rarement détectable au moyen du témoignage de l'introspection [*introspective evidence*]. L'effet de boire du lait difficilement digérable pourrait troubler davantage l'estomac de Lonnie et aggraver sa déshydratation. En revanche, Lonnie-Plus peut voir que l'ingestion d'une abondance de liquides clairs pourrait rapidement améliorer la condition physique de Lonnie – ce qui, incidemment, pourrait également l'aider en ce qui concerne son mal du pays. Lonnie-Plus peut également voir à quel point le fait de boire des liquides clairs serait désagréable pour Lonnie; ce qui arriverait précisément si Lonnie continuait à souffrir de déshydratation, et ainsi de suite. Sur la base de ces informations, Lonnie-Plus pourrait alors en venir à désirer tout cela s'il devait prendre la place de Lonnie; il voudrait boire des liquides clairs plutôt que du lait, ou du moins, agir de telle façon qu'un désir de ce type soit satisfait. La base réductive [*reduction basis*] de cet intérêt objectifié implique des faits au sujet de la constitution et des circonstances de Lonnie, lesquels déterminent, entre autres, ses goûts actuels et sa capacité de développer certains nouveaux goûts, les conséquences occasionnées par la déshydratation prolongée, les effets et la disponibilité de plusieurs types de liquides, et ainsi de suite.

Affirmons, par conséquent, que cette base réductive constitue la constellation des qualités premières qui fait en sorte que Lonnie a un certain *intérêt objectif*.¹⁶ C'est pourquoi nous pourrions affirmer que Lonnie a un intérêt objectif de boire des liquides clairs en vertu de cet ensemble de faits complexes, relationnels et dispositionnels. Autrement dit, nous pourrions affirmer que la base réductive, et non pas le simple fait que Lonnie-Plus aurait certains désirs [*wants*], est le vérificateur [*truth-maker*] de l'idée qu'il s'agit là d'un intérêt objectif de Lonnie. L'intérêt objectif explique ainsi la raison pour laquelle il existe un certain intérêt objectifié, et non pas l'inverse.¹⁷

Affirmons à présent que *X* est *non moralement bon* [good] pour *A* si et seulement si *X* satisfaisait un intérêt objectif de *A*.¹⁸ Nous pourrions penser au sujet des croyances de *A*⁺ concernant ce qu'il voudrait vouloir s'il était à la place de *A*, qu'elles pourraient donner lieu à un classement des intérêts objectifs possibles de *A*, un classement qui pourrait refléter ce qui est le mieux ou le pire pour *A*, et qui nous permettrait de parler des désirs [wants] actuels de *A* en tant qu'il s'agit de meilleures ou de pires approximations de ce qui est le mieux pour lui. Nous pourrions aussi décomposer le point de vue de *A*⁺ en des intérêts objectifs *prima facie* par opposition à des intérêts objectifs « tout bien considéré » [on balance] de *A*, où ces derniers se rapportent à la notion d'« un bien pour *A* » [a good for *A*] alors que les premiers renvoient à la notion « du bien de *A* » [the good for *A*].¹⁹

Il me semble qu'il s'agit là d'une explication plausible de ce en quoi consiste le bien [good] non moral de quelqu'un : grosso modo, ce qu'il voudrait lui-même convoiter s'il savait ce qu'il fait.²⁰

Qui plus est, cette explication conserve ce qui, selon moi, semble être une relation adéquate entre les valeurs non morales et la motivation. Supposons que l'on désire *X*, mais que nous nous demandions si *X* fait réellement partie de son bien [good]. Ce sentiment de perplexité survient généralement parce que l'on croit que l'on ne sait pas assez de choses sur *X*, sur soi ou sur son monde, ou bien parce qu'on sent que l'on n'est pas adéquatement rationnel ou réflexif [reflective] en évaluant l'information dont on dispose – on soupçonne peut-être que l'on a été captivé par quelques caractéristiques saillantes de *X* (ou bien repoussés par quelques caractéristiques saillantes d'options autres que *X*). Si l'on devait apprendre que l'on continuerait néanmoins à vouloir pour soi-même *X* dans des circonstances où l'on pourrait considérer les choses de manière à ce que l'on soit pleinement informé et rationnel, cela donnerait vraisemblablement moins de poids à l'inquiétude initiale. En revanche, si l'on devait apprendre que, étant pleinement informé et rationnel, on *ne* voudrait *pas* vouloir pour soi-même *X* dans ces circonstances, cela donnerait vraisemblablement plus de poids à cette inquiétude. Les désirs étant ce qu'ils sont, une inquiétude renforcée pourrait ne pas suffire pour éliminer le désir de *X*. Mais si l'on devait être sincèrement et vivement convaincu que le fait de désirer *X* n'est, en ce sens, pas appuyé par une réflexion exhaustive sur les faits, on pourrait vraisemblablement ressentir qu'il s'agit là d'une raison pour ne pas agir conformément à ce désir. L'ajustement du désir à la croyance ne saurait, dans un cas donné, être requis par la raison ou la logique; cet ajustement pourrait être « tout simplement [merely] psychologique ». Mais ce sont précisément des phénomènes psychologiques de ce type que les théories naturalistes de la valeur considèrent comme fondamentaux [basic].

Dans ce qui suit, nous devons avoir recours à la notion de « bien intrinsèque » [intrinsic goodness]. Nous affirmerons alors que *X* est un bien intrinsèquement non moral [intrinsically non-morally good] de *A* seulement si *X* constitue un intérêt objectif de *A* et n'implique aucune référence à tout autre intérêt objectif de *A*. Nous pouvons employer en un sens évident la notion d'intérêt intrinsèque

objectif afin de rendre compte de tout autre intérêt objectif. Étant donné que les individus et les environnements diffèrent à bien des égards, il n'est pas nécessaire de présupposer qu'ils ont tous les mêmes intérêts intrinsèques objectifs. À fortiori, nous n'avons pas besoin de présupposer qu'ils sont caractérisés par les mêmes intérêts instrumentaux objectifs. Nous devrions, toutefois, nous attendre à ce qu'il existe, à cet égard, des similarités correspondantes dans leurs intérêts lorsqu'il existe des similarités personnelles et situationnelles entre les individus, c'est-à-dire lorsqu'il existe des similarités dans les bases réductives.

Il est désormais possible de voir la façon dont la notion de bien [*goodness*] non moral se prête à des usages explicatifs. Pour commencer, elle peut nous permettre d'expliquer pourquoi les désirs actuels de quelqu'un peuvent avoir certaines caractéristiques contrefactuelles – par exemple, pourquoi quelqu'un aurait certains désirs hypothétiques plutôt que d'autres s'il devait être pleinement informé et conscient. Un usage explicatif de ce type – qui découle directement de la définition d'intérêt objectif – pourrait toutefois bien être jugé décevant à moins que d'autres fonctions explicatives soient trouvées.

Considérons alors la différence entre Lonnie et Tad, un autre voyageur se trouvant dans les mêmes lieux, mais qui, à la différence de Lonnie, désire boire des liquides clairs et agit conformément à ce désir. Tad sera ravivé alors que Lonnie demeurera léthargique. Nous pourrions expliquer cette différence en soulignant que, bien que Lonnie et Tad aient tous les deux agi en fonction de leurs désirs [*wants*], les désirs [*wants*] de Tad reflètent mieux ses intérêts. La congruence entre les désirs [*wants*] de Tad et ses intérêts peut être fortuite, tout comme il est possible que Tad soit au courant de sa déshydratation et connaisse le traitement standard à suivre. Dans ce dernier cas, nous affirmerions en général que l'explication de la différence de leurs conditions tient au fait que Tad, et non Lonnie, « savait ce qui était bien [*good*] pour lui. »

Nous pouvons anticiper que, en général, ce que A^+ voudrait vouloir s'il était à la place de A corrèlerait bien avec ce qui permettrait à A d'éprouver un état de bien-être physique ou psychologique, et d'échapper à un état de malaise physique ou psychologique. Bien sûr, notre état de bien-être ou de malaise figure parmi les choses qui nous importent le plus, et ce, de manière la plus fiable, même après mûre réflexion.²¹ Le recours à des degrés de congruence entre les désirs [*wants*] et les intérêts de A nous aidera souvent à expliquer les faits à propos du degré de satisfaction de A concernant sa propre vie. L'explication ne pourrait être préservée si l'on substituait « le fait de croire qu'il y a congruence » avec « le fait d'être congruent (à un tel ou tel degré) », étant donné que, comme le démontrent des cas comme celui de Lonnie, même si nous parvenions à nous convaincre nous-mêmes que nos désirs [*wants*] reflètent nos intérêts, le fait d'agir conformément à ces désirs [*wants*] pourrait ne pas engendrer beaucoup de satisfaction.

En vertu de la corrélation escomptée entre le fait d'agir en fonction de motifs qui concordent avec nos propres intérêts et le fait de parvenir à un certain niveau de

satisfaction et d'éviter la détresse, nos intérêts objectifs peuvent aussi jouer un rôle explicatif dans l'évolution de nos désirs. Considérons ce que j'appelle le *mécanisme désirs* [wants]-*intérêts*, lequel permet aux individus d'apprendre, sur la base de leurs expériences, ce en quoi consistent leurs intérêts dans un état à la fois de conscience-de-soi [*selfconscious*] et de non-conscience-de-soi [*unself-conscious*]. Dans les cas les plus simples, l'apprentissage par essais et erreurs conduira à une rétention sélective de désirs [wants] qu'il est possible de satisfaire [*satisfiable*] et à des résultats satisfaisants pour l'agent.

Supposons, par exemple, que Lonnie cède à son envie de boire le lait. Peu de temps après, son état se dégrade. Étant toujours incapable d'identifier la source de son malaise et resté aux prises avec le désir de ce qui est familier, son attention se tourne vers un signe vert et rouge dans la fenêtre d'un petit commerce sur son passage, indiquant « 7-Up ». Il se précipite à l'intérieur et achète une bouteille. Bien que tiède, il la boit avec impatience. « Mmm », pense-t-il, « j'en prendrais une autre ». Il achète une seconde bouteille et la boit. À ce moment, sa soif de soda tiède est étanchée et il poursuit son chemin. Après quelques heures, son humeur s'améliore. Lorsqu'il passe de nouveau devant le commerce sur le chemin vers son hôtel, l'association plaisante liée au fait de boire du 7-Up l'incite à en acheter de nouveau et à l'emporter avec lui. Ce soir-là, dans la sombre solitude de sa chambre, il retrouve le goût familier et rassurant du soda réconfortant, et boit quelques bouteilles de plus avant de finalement s'endormir. Au moment de son réveil le lendemain matin, il se sent beaucoup mieux. Pour résumer une histoire ennuyeuse : la prochaine fois que Lonnie se sentira mal en point à l'étranger, il pourrait avoir une certaine tendance consciente ou inconsciente, raisonnée ou superstitieuse d'être à la recherche du 7-Up. S'il n'est pas en mesure d'en trouver, il pourrait alors se mettre à la recherche d'un soda à saveur de lime, spécifique au pays où il se trouve, ou peut-être même l'*agua mineral con gaz* qu'il méprisait autrefois. Peu à peu, au fur et à mesure que Lonnie voyage davantage et souffre d'un malaise semblable, il en viendra à boire régulièrement des liquides clairs et, de manière régulière, à se sentir mieux, de sorte qu'il développera finalement un désir actuel pour de tels liquides – et une aversion pour d'autres liquides, tels que le lait – dans de telles circonstances.

C'est ainsi que les désirs de Lonnie ont évolué au cours de son expérience pour se conformer de plus près à ce qui est bien pour lui, au sens naturaliste du terme. Le processus ne relève pas d'une réponse idéalement rationnelle au fait de recevoir des informations idéales, mais plutôt d'une expérimentation en grande partie non réfléchie, accompagnée d'associations et de renforcements positifs et négatifs. Il n'y a aucune garantie que les désirs « appris » par de la rétroaction de ce type reflètent précisément ou complètement le bien [*good*] de l'individu. Il est encore moins certain que, même lorsqu'un ajustement adéquat dans les désirs se produit, l'agent comprenne l'origine de ses désirs ou soit capable de se représenter la nature des intérêts qu'ils reflètent. Mais il s'agit alors d'une caractéristique tout à fait générale propre aux divers moyens par lesquels nous apprenons à connaître le monde, qu'ils puissent ne pas fournir des représentations précises [*accurate*] ou complètes de ce monde. Ma capacité de percevoir

et de comprendre mon environnement coexiste avec, voire s'appuie sur les mêmes mécanismes que ma disposition [*liability*] à être sujet à la duperie par la voie d'illusions, d'attentes ou d'apparences.

Il existe des raisons largement théoriques de penser qu'il existe quelque chose comme un mécanisme désirs [*wants*]-intérêts et qu'il a un rôle important dans la formation de désirs. Les êtres humains sont des créatures principalement motivées par des désirs plutôt que par des instincts. Si de telles créatures étaient incapables au cours de leurs expériences d'adapter leurs désirs [*wants*] de manière plus étroite à leurs intérêts essentiels – peut-être parce qu'elles ne sont pas plus susceptibles de faire l'expérience d'états internes positifs lorsque leurs intérêts essentiels sont servis plutôt que lorsqu'ils ne le sont pas –, nous serions portés à croire qu'elles ne connaîtraient pas un long et fructueux avenir. Ainsi, si les êtres humains en général n'en étaient pas venus à manger le type de nourriture qui leur assure un certain degré de bien-être physique, ou à s'engager dans le type d'activités ou de relations qui leur permet de maintenir une bonne santé mentale, nous ne serions pas ici aujourd'hui en train de nous interroger sur ce qui est bon [*good*] pour nous. Étant donné que des créatures aussi sophistiquées et complexes que les êtres humains ont évolué dans des contextes liés à une diversité d'environnements et ont, par ailleurs, pris l'habitude de modifier leurs environnements, nous devrions alors nous attendre à ce que nous fassions preuve d'une flexibilité considérable quant à notre capacité d'adapter, au cours de notre expérience, nos désirs [*wants*] à nos intérêts. Toutefois, cette même flexibilité fait en sorte qu'il s'agit là d'un mécanisme qui n'est pas fiable [*unreliable*] : nos désirs [*wants*] peuvent, à n'importe quel moment, considérablement différer de nos intérêts, et ce, de manière arbitraire. En outre, nous pourrions ne pas être en mesure d'avoir des expériences qui pourraient nous amener à remarquer ceci, ou de faire suffisamment l'objet de rétroaction pour avoir la chance de développer de nouveaux désirs [*wants*] qui se rapprochent davantage de nos intérêts. Il est tout à fait possible, et rarement peu fréquent, qu'un individu vive le reste de sa vie normale sans jamais reconnaître certains de ses intérêts les plus fondamentaux ou s'ajuster à eux. Les mécanismes culturels d'acquisition de désirs [*wants*] peuvent en partie remédier aux limitations individuelles, et favoriser l'apprentissage et même la théorisation sur l'étendue de plusieurs vies et durées de vie, mais ces mêmes mécanismes créent aussi un vaste potentiel d'inculcation de désirs [*wants*] qui vont à l'encontre d'intérêts.

On peut conférer à la thèse selon laquelle il existe un mécanisme désirs [*wants*]-intérêts relativement le même statut et la même plausibilité décontractée [*breezy*] que l'argument plus strictement biologique selon lequel nous devons nous attendre à ce que l'œil humain soit capable de détecter des objets d'une taille et d'une forme similaires à celles de nos prédateurs ou de nos proies. Il n'est pas nécessaire de ne supposer rien de tel qui corresponde plus ou moins à l'infailibilité. Il nous suffit de présupposer assez de succès fonctionnel, en vertu duquel nous parvenons à tenir notre rang dans un monde très souvent inhospitalier.²²

Jusqu'ici, l'argumentaire n'a porté que sur les intérêts objectifs qui peuvent être caractérisés en tant que besoins, mais le mécanisme désirs-intérêts peut s'appliquer à n'importe quel intérêt – même les intérêts liés aux aptitudes particulières ou aux rôles sociaux d'un quelconque individu – dont la frustration s'accompagne – et ce, même indirectement – de résultats consciemment ou inconsciemment insatisfaisants pour cet individu. (Bien entendu, plus l'association est indirecte, plus il est improbable que le mécanisme soit fiable.) Par exemple, l'expérience consistant à suivre des cours de mathématiques et de philosophie pourrait inciter une étudiante qui se croyait jusque-là destinée à une carrière de mathématicienne à en venir à préférer une carrière de philosophe, laquelle serait en fait mieux adaptée à ses aptitudes et à ses attitudes. Un ouvrier récemment promu à un poste de gestionnaire pourrait se trouver moins enclin à répondre aux griefs des employés – ce qu'il revendiquait pourtant auparavant auprès de ses gestionnaires –, alors que ses anciens collègues pourraient, de même, se trouver moins enclins qu'auparavant à se confier à lui.

Si un mécanisme désirs-intérêts est postulé, et si ce qui est non moralement bien pour un individu est une question qui relève de son intérêt objectif, nous pouvons alors affirmer que les valeurs objectives sont à même de jouer un rôle dans l'explication des valeurs subjectives du type requis par le réalisme naturaliste des valeurs. Ces explications fournissent même un appui à certaines prédictions qualifiées [*qualified predictions*] : par exemple, que, toutes choses étant égales par ailleurs, les individus seront habituellement les mieux placés que de tierces parties pour juger de leurs propres intérêts; que des individus possédant des caractéristiques personnelles et sociales similaires tendront à avoir des valeurs similaires; et qu'un consensus général plus important pourrait être dégagé sur ce qui est désirable dans les domaines de la vie où les individus sont semblables à d'autres égards (par exemple, sur le plan des motivations de base) et où l'on prévoit que les mécanismes d'essais et erreurs fonctionnent bien (par exemple, lorsqu'un savoir ésotérique n'est pas requis). Je ne suis pas en mesure de me prononcer sur l'exactitude de ces prédictions, mais celles-ci concordent avec des opinions largement répandues, ce qui est tout à son honneur.

Il conviendrait peut-être de souligner que, bien que je parle d'objectivité des valeurs, les valeurs en question sont humaines et existent seulement parce que les êtres humains existent. Du point de vue d'une théorie de la valeur classique [*old-fashioned*], il s'agit d'une notion relationnelle – plutôt qu'absolue – de ce qui est bon [*goodness*]. Bien qu'ils soient relationnels, les faits pertinents au sujet des êtres humains et de leur monde sont objectifs de la même façon dont les entités non relationnelles, telles que les pierres, le sont : leur existence ou leur nature ne dépend pas simplement de la conception que nous en avons.²³

Compris ainsi, les intérêts objectifs surviennent sur les faits naturels et sociaux. Est-ce que cela signifie que les intérêts objectifs ne peuvent avoir aucune valeur explicative, étant donné qu'il devrait toujours être possible, en principe, de rendre compte de n'importe quel fait particulier qu'ils prétendent expliquer par référence seulement à la base de survenance? Si la survenance était à elle seule

une raison de refuser un rôle explicatif à une série de concepts quelconques, nous devrions alors affirmer que la chimie, la biologie et le génie électrique – lesquels surviennent manifestement sur la physique – sont dépourvus de pouvoir explicatif. En effet, même la possibilité d'une réduction catégorique n'est pas une raison de douter de ce pouvoir explicatif. Le fait d'établir une relation de réduction entre, par exemple, un phénomène chimique tel que la valence et un modèle physique de l'atome ne suggère aucunement qu'il n'y a pas de phénomène tel que la valence, ou que les généralisations se référant à la valence ne peuvent fournir un appui à des explications. Il ne peut être question ici d'économie ontologique ou de rejeter des entités non nécessaires, tel que ce serait le cas si, par exemple, la valence était effectivement considérée comme un phénomène *sui generis*, au-delà de toute constellation de propriétés physiques. Les faits décrits au moyen des principes de la valence chimique sont authentiques, et ils permettent d'établir une systématisation puissante et explicative des combinaisons chimiques. Qu'une réduction de la chimie à la physique atomique ait été accomplie ne fait que renforcer ces affirmations.

Nous sommes persuadés que la notion de valence chimique est explicative puisque les explications formulées en termes de valence chimique inscrivent des *explananda* dans un cadre nomique distinct et bien articulé, de manière à ce qu'elle en augmente notre compréhension de manière évidente. Mais existe-t-il une théorie aussi puissante et explicative de la notion d'intérêt objectif, qui peut nous donner des raisons de penser – peu importe si une réduction stricte est possible ou non – que les explications élaborées au moyen de cette notion sont véritablement informatives?

Le type de réalisme des valeurs esquissé ici m'apparaîtrait inintéressant s'il me semblait qu'aucune théorie d'importance ne pourrait être développée au moyen de la catégorie de « valeur objective ». Mais en décrivant le mécanisme désirs-intérêts, j'ai déjà tenté d'indiquer qu'une théorie de ce genre est possible. Lorsque nous cherchons à expliquer les raisons pour lesquelles ils agissent ainsi, les raisons pour lesquelles ces individus ont certaines valeurs ou certains désirs, et ce qui parfois les conduit au conflit et, à d'autres moments, à la coopération, il est tout fait naturel, du point de vue du sens commun de même que des sciences sociales, de parler en termes d'intérêts individuels. De telles explications demeureront pourtant incomplètes et superficielles tant et aussi longtemps que nous nous en tiendrons seulement aux intérêts subjectifs, puisque ceux-ci doivent également faire l'objet d'une explication.²⁴

IV. LE RÉALISME NORMATIF

Supposons que tout ce que j'ai affirmé jusqu'à maintenant est généreusement tenu pour acquis. Je n'aurais pourtant pas encore le droit de parler de réalisme *moral* puisque je n'aurais fait valoir que la possibilité d'un type de réalisme à propos de ce qui est non moralement bon [*non-moral goodness*], une notion pouvant même être admise par de parfaits sceptiques en matière de moralité. Afin de pouvoir parler de réalisme moral, je dois montrer comment le réalisme

des valeurs typiquement morales ou des normes morales est possible. Je me concentrerai sur les normes morales, c'est-à-dire sur ce qui a trait au bien et au mal moral [*matters of moral rightness and wrongness*]; mais l'argument que je présente peut, par extension, s'appliquer également aux valeurs morales. Ce choix s'explique, en partie, par le fait que le réalisme normatif semble intuitivement beaucoup moins plausible que le réalisme des valeurs. Il n'est donc pas surprenant que plusieurs propositions actuelles mises de l'avant en faveur du réalisme moral se concentrent essentiellement sur la question des valeurs – et parfois seulement sur ce qui constitue des valeurs non morales. Toutefois, peu importe la conception de la moralité proposée, une théorie morale doit présenter une conception du bien [*rightness*].

Le réalisme moral normatif n'est pas plausible pour diverses raisons, mais dans le cadre du présent essai, la raison plus pertinente tient au fait qu'il me semble impossible d'étendre la stratégie générique du réalisme naturaliste aux normes morales. Quelle place revient aux faits concernant ce qui *doit* être le cas dans le cadre d'une explication? Est-ce que les faits concernant la façon dont les choses *sont* dans le monde [*the way things are*] suffisent pour expliquer la façon dont on doit agir? Bien sûr qu'il en est ainsi. Mais alors, ma conception naturaliste du réalisme moral m'engage à considérer les faits au sujet de ce qui doit être le cas comme une espèce particulière de faits concernant la façon dont les choses sont dans le monde. Par conséquent, il est possible que ces faits puissent avoir une fonction dans le cadre d'une théorie explicative. Pour envisager comment cela peut être possible, je propose que l'on considère quelques exemples d'explications impliquant des normes naturalisées en dehors du domaine de la moralité.

« Pourquoi le toit s'est-il effondré? – Pour qu'une maison puisse supporter des charges de neige de cette teneur, la charpente doit être constituée de planches de bois qui mesurent au moins 2 pouces x 8 pouces, et non pas 2 pouces x 6 pouces ». Cette explication est tout à fait acceptable en soi, mais elle implique un « doit » [*ought*]. Nous pourrions, bien sûr, éliminer ce « doit » [*ought*] comme suit : « s'il est pour être à même de résister à de telles chutes de neige, un toit conçu en fonction de ce modèle est alors dans l'obligation de disposer d'une charpente avec une coupe transversale d'au moins 2 pouces x 8 pouces ». Au « doit » [*ought*] architectural se substitue un « si... alors... » qui relève de l'ingénierie. Cela est possible puisque le « doit » est clairement hypothétique, ce qui reflète l'objectif architectural, universel consistant à fabriquer des toits suffisamment forts pour qu'ils ne s'effondrent pas. Puisque l'objectif est fixé selon le contexte, qu'il y a des réponses plus ou moins définies apportées à la question de savoir comment cet objectif peut être réalisé et, en outre, que le phénomène à expliquer est le résultat d'un processus qui implique une sélection contre les instances qui n'atteignent pas cet objectif, l'explication qui implique le « doit » transmet de l'information explicative.²⁵ Je qualifierai ce type d'explication de « *critériel* » : nous expliquons pourquoi quelque chose se produit par référence à un critère pertinent, étant donné l'existence d'un processus qui sélectionne en faveur (ou contre) des phénomènes qui se rapprochent approximativement de ce critère. Bien que le critère soit défini sous l'angle naturaliste, il pourrait en

même temps être de nature à jouer un rôle régulateur dans le cadre des pratiques humaines; en l'occurrence, la construction de bâtiments.

Un type plus familier d'explication critériale concerne les normes de rationalité individuelle. Considérons le recours à une théorie instrumentale de la rationalité afin d'expliquer le comportement d'un individu à la lumière de ses croyances et de ses désirs, ou de rendre compte de la manière dont ses croyances se modifient avec l'expérience.²⁶ Bobby Shaftoe est parti en mer parce qu'il croyait que c'était la meilleure façon de faire fortune, et il voulait avant tout faire fortune. Son coéquipier Reuben Ramsøe en est venu à la conclusion qu'il n'était pas apprécié par les autres matelots puisqu'il constatait qu'ils se moquaient de lui et ne dissimulaient pas leur appréciation pour les flagellations constantes qui lui étaient infligées par le second capitaine. La validité de ces explications tient au fait que l'action ou la croyance en question était tout à fait rationnelle du point de vue de l'agent qui se trouvait dans ces circonstances, et parce que nous supposons correctement que Shaftoe et Ramsøe devaient être tout à fait rationnels.

Les faits portant sur les degrés de rationalité instrumentale s'inscrivent dans le cadre d'explications par d'autres moyens également. Considérons d'abord la question de savoir pourquoi Bobby Shaftoe a connu plus de succès que la plupart de ses semblables dans la réalisation de ses buts. Nous pourrions attribuer son succès au fait que Shaftoe est mieux doté de raison instrumentale que la plupart des individus – peut-être possède-t-il une perspicacité supérieure à la moyenne pour estimer la probabilité des conséquences [*outcomes*], ou bien fait-il preuve d'une fiabilité supérieure à la moyenne en matière d'inférences déductives, ou dispose-t-il d'une imagination supérieure à la moyenne pour évaluer diverses options.

En deuxième lieu, bien que nous soyons tous imparfaits en matière de délibération, notre comportement peut en venir à incarner des habitudes et des stratégies qui nous permettent de nous rapprocher d'une rationalité optimale de manière plus étroite que ce que nos défauts de délibération [*deliberative defects*] pourraient nous faire croire. Le mécanisme est simple. Des schémas [*patterns*] de croyances et de comportements qui manifestent une faible rationalité instrumentale tendent à être, dans une certaine mesure, voués à l'échec [*self-defeating*]; une incitation qui favorise leur modification. En revanche, les schémas qui manifestent une forte rationalité instrumentale tendent à être, dans une certaine mesure, gratifiants [*rewarding*]; une incitation qui favorise leur reproduction. Ces incitations peuvent influencer nos croyances et nos comportements même si les inconvénients ou les avantages des schémas en question ne font pas l'objet d'une délibération consciente. Dans de tels cas, nous pourrions affirmer que nous acquérons ces habitudes ou ces stratégies parce qu'ils *sont* plus rationnels, sans qu'il y ait une quelconque *croyance* intermédiaire qui donne à penser cela. Les psychologues cognitifs ont ainsi exposé certaines des stratégies ou heuristiques inconscientes que nous employons afin de permettre à nos intellects limités de trier davantage de données et de porter des jugements plus rapides et plus constants que ce serait le cas si nous avions seulement recours à

des formes plus standards de raisonnement explicite.²⁷ À notre insu, nous en venons à dépendre d'heuristiques en partie parce qu'elles sont sélectivement renforcées par leurs avantages instrumentaux par rapport au raisonnement standard et explicite, c'est-à-dire en raison de leur plus grande rationalité. De même, nous pouvons – sans que nous le réalisions ou nous puissions l'admettre à nous-mêmes – développer des schémas de comportements qui encouragent ou découragent des comportements particuliers chez autrui, tels que le mécanisme inconscient par lequel nous causons chez les individus dont nous n'apprécions pas la compagnie le fait de ne pas apprécier la nôtre. Finalement, en tant qu'enfants, nous étions pratiquement incapables de procéder à des évaluations rationnelles lorsque l'acquisition d'un gain éloigné dans le temps impliquait une perte imminente. Pourtant, d'une certaine manière, nous sommes parvenus de manière largement non délibérative, avec le temps, à acquérir diverses habitudes intéressantes, telles que la capacité de déplacer vers la périphérie de notre attention certaines pensées vives au sujet de l'avenir immédiat, capacité qui nous permet en tant qu'adultes de nous rendre par nous-mêmes chez le dentiste sans être poussés par derrière. L'explication critérielle en termes de rationalité individuelle s'étend ainsi aux comportements bien au-delà de l'action délibérée. Il devient alors possible de voir dans l'émergence de comportements de ce genre, tout comme dans le mécanisme désirs-intérêts, quelque chose que nous pourrions nommer sans déformation « apprentissage ».

En effet, notre tendance à développer au fur et à mesure des habitudes et des stratégies rationnelles peut coopérer avec le mécanisme désirs-intérêts afin de servir de base à une forme *étendue* d'explication critérielle, où la rationalité d'un individu est évaluée non pas en fonction de ses croyances et de ses désirs actuels [*occurrent*], mais en fonction de ses intérêts objectifs. Les exemples concernant le mécanisme désirs-intérêts que nous avons considérés antérieurement impliquent des éléments d'une explication de ce type, en ce qu'ils démontrent non seulement que les désirs sont ajustés en fonction des intérêts, mais aussi que le comportement est ajusté en fonction de désirs nouvellement ajustés. Sans la présence d'une modification adéquate du comportement, qui reflète les désirs en évolution, la rétroaction nécessaire à notre prise de connaissance de nos désirs ne se produirait pas. Avec une modification de ce type, le comportement peut lui aussi devenir plus rationnel au sens élargi du terme. Un individu qui fait preuve de rationalité instrumentale peut être disposé à ajuster ses moyens en fonction de ses fins, mais ce qui peut résulter du fait d'avoir recours à un moyen particulier – choisir un cours, accepter un nouvel emploi – peut être une évaluation plus informée de ses fins et peut-être une reconsidération de celles-ci.

La théorie de la rationalité individuelle – que ce soit dans sa forme simple ou étendue – nous offre ainsi une instance du type requis pour l'élaboration d'un exemple de réalisme normatif. Les évaluations des degrés de rationalité instrumentale jouent un rôle majeur dans nos explications du comportement des individus, mais elles fonctionnent aussi simultanément à titre de force normative pour l'agent. Peu importe les autres soucis que peut avoir un agent, il s'agit sûrement, pour elle ou lui, d'une caractéristique positive de l'action que l'efficacité

de celle-ci soit relative à ses croyances ou à ses désirs ou, dans un sens étendu, relative aux croyances et aux désirs qui reflètent adéquatement ses conditions et ses circonstances.

La force normative manifestée par ces théories de la rationalité individuelle ne doit pas, toutefois, être simplement dérivée de leur usage explicatif. On peut employer une théorie de la rationalité instrumentale afin d'expliquer des comportements et la rejeter à titre de théorie normative des raisons, tout comme on peut expliquer une action par son irrationalité sans pour autant alors promouvoir la déraison.²⁸ La relation entre les rôles normatifs et explicatifs de la conception instrumentale de la rationalité doit plutôt être reconduite à leur terrain commun : le système motivationnel humain. Il s'agit d'un fait, en ce qui nous concerne, que nous avons à la fois des fins et la capacité d'entreprendre des actions délibérées en fonction de celles-ci et d'ajuster non délibérément notre comportement relativement à nos fins. Par conséquent, des options s'offrent sur des chemins de divers horizons de possibilités que nous évaluons différemment. Lorsque nous expliquons les raisons à l'origine des choix des personnes et les causes de leur comportement, de même que lorsque nous faisons appel à leurs intuitions quant à la rationalité de leurs décisions ou de leurs actions éventuelles, nous œuvrons sur ce terrain puisque nous faisons appel dans la mesure du possible aux faits concernant ce qui motive les agents dans les faits ou en principe.

C'est ainsi qu'émerge la possibilité d'affirmer que des faits existent au sujet de ce que les individus ont des raisons de faire, des faits qui peuvent être considérablement indépendants de – et plus probants normativement que – l'idée que se fait actuellement l'agent de ses raisons. L'argument en faveur d'un tel réalisme à propos de la rationalité individuelle n'est pas plus probant que les arguments soutenant la double proposition selon laquelle la conception de la rationalité instrumentale et individuelle en question est caractérisée à la fois par un pouvoir explicatif et par un type de force approbative qu'une théorie des *raisons* se doit de posséder, mais ces arguments (bien que je n'en discuterai pas davantage ici) semblent, selon moi, très convaincants.

* * *

Procédons maintenant au-delà de la théorie de la rationalité individuelle et demandons-nous à quoi pourraient ressembler des explications critérielles [*criterial explanations*] qui impliquent des normes typiquement morales. Afin de répondre à cette question, nous devons savoir ce qui distingue une norme morale d'autres critères d'évaluation. L'évaluation morale semble porter avant tout sur l'évaluation de conduites ou de caractères où les intérêts de plus d'une personne sont en jeu. Qui plus est, l'évaluation morale évalue les actions ou les conséquences [*outcomes*] de manière singulière : les intérêts du plus fort ou du plus prestigieux ne prévalent pas toujours, des raisons purement prudentielles peuvent être subordonnées, et ainsi de suite. Plus généralement, les résolutions morales sont considérées comme déterminées par des critères de choix qui sont *non indexicaux* et, en un certain sens, *englobants* [*comprehensive*]. Cela a incité un

certain nombre de philosophes à vouloir rendre compte du caractère spécifique de l'évaluation morale en identifiant un *point de vue moral* qui est impartial, mais qui prend également en considération tous les individus pouvant être éventuellement concernés. D'autres théoriciens en éthique en sont venus à une conclusion similaire en examinant les types de raisons que nous considérons typiquement comme pertinents ou non pertinents dans le cadre du discours moral. Poursuivons sur ces voies. Nous pouvons ainsi affirmer que les normes morales reflètent un certain type de rationalité, une rationalité qui relève non pas du point de vue d'un individu en particulier, mais de ce que nous pourrions appeler un point de vue social.²⁹

En elle-même, l'adéquation [*equation*] entre le bien moral [*moral rightness*] et la rationalité d'un point de vue social n'est pas extrêmement restrictive, car selon ce que l'on entend par rationalité, cette adéquation pourrait être établie par un utilitariste, un kantien ou même un non cognitiviste. Il doit en être ainsi puisque si cette adéquation doit pouvoir saisir ce qui est spécifique aux normes morales, elle doit être compatible avec le plus vaste éventail de théories morales reconnues. Toutefois, dès que l'on opte pour une conception particulière de la rationalité – telle que la conception de la rationalité conçue comme la recherche efficiente du bien non moral [*the non-morally good*], ou comme autolégislation autonome et universelle, ou comme l'expression non cognitive d'une approbation hypothétique [*hypothetical endorsement*] – cette caractérisation schématique commence alors à assumer un contenu moral. Nous avons adopté ici la conception instrumentale de la rationalité, et cela – en lien avec la conception de ce qui est non moralement bon (*non-moral goodness*) – signifie que l'argument en faveur du réalisme moral présenté ci-dessus consiste en un argument qui présuppose et prétend défendre une théorie morale particulière et substantielle.³⁰

En quoi consiste cette théorie? Introduisons d'abord une idéalisation de la notion de rationalité sociale en considérant ce qui serait rationnellement approuvé si tous les intérêts de tous les individus éventuellement concernés étaient pris en considération dans des circonstances où l'on dispose d'informations vives et complètes.³¹ Étant donné la supposition suivant laquelle nous disposons d'informations vives et complètes, les intérêts en question seront objectifs. Considérant la théorie du bien [*goodness*] proposée dans la section III, cette idéalisation correspond à ce qui est rationnel d'un point de vue social, relativement à la réalisation du bien intrinsèque non moral [*intrinsic non-moral goodness*]. Il semble, selon moi, qu'il s'agit là d'un critère reconnaissable et plausible, voire très peu controversé du bien moral [*moral rightness*]. Le bien moral relatif [*relative moral rightness*] consiste en un certain degré d'approximation de ce critère.

La question se pose maintenant de savoir si la notion de degrés de bien moral [*moral rightness*] peut intervenir dans les explications des comportements ou des processus d'apprentissage moral qui existent parallèlement aux usages explicatifs que permet la notion de degrés de rationalité individuelle, comprise, en particulier, en un sens étendu. Je tenterai de suggérer divers moyens qui rendraient possibles de tels usages.

Tout comme l'individu qui néglige de manière considérable certains de ses intérêts sera susceptible [*liable*] d'éprouver diverses formes d'insatisfaction, de même une organisation sociale [*social arrangement*] – par exemple, une forme de production, une hiérarchie sociale ou politique, etc. – qui dévie par rapport à la rationalité sociale en négligeant considérablement les intérêts d'un groupe spécifique sera potentiellement sujette à l'insatisfaction et à l'agitation. Que ce potentiel soit réalisé dépend d'un ensemble considérable de circonstances. En raison de la socialisation, ou d'autres limitations s'exerçant sur l'expérience ou la connaissance des membres de ce groupe, le mécanisme désirs-intérêts ne pourrait pas avoir fonctionné de manière à ce que les désirs de ses membres reflètent leurs intérêts. Par conséquent, les individus peuvent n'éprouver aucune frustration directe de leurs désirs malgré la mise à l'écart de leurs intérêts. Le groupe peut également être trop dispersé ou trop faible pour se mobiliser de manière efficace. Il se peut également qu'il fasse l'objet d'une répression sévère. D'autre part, certaines circonstances sociales et historiques favorisent la réalisation de ce potentiel d'agitation sociale, telles que, par exemple, le fait de fournir aux membres de ce groupe des expériences qui ont pour conséquence de les rendre plus susceptibles de développer des intérêts s'arrimant à leurs désirs, d'affaiblir l'appareil de répression actuel, de fournir aux membres de ce groupe un accès inédit à des ressources ou à de nouvelles opportunités de mobilisation, ou simplement de dissiper l'illusion que le changement est impossible. On peut s'attendre à ce que le potentiel d'agitation sociale se manifeste alors dans de telles circonstances.

Tout comme les explications impliquant des évaluations de la rationalité individuelle ne peuvent être remplacées par des explications impliquant des *croyances individuelles à propos* de ce qui doit être rationnel, de même les explications impliquant des évaluations au sujet de la rationalité sociale ne peuvent être remplacées par des *croyances à propos* de ce qui est moralement bien [*right*]. Par exemple, un sentiment de mécontentement peut surgir lorsqu'une société dévie par rapport à la rationalité sociale, mais non pas en conséquence de la croyance que cela est effectivement le cas. Supposons que tous les membres d'une société quelconque croient que celle-ci est juste. Cette croyance peut contribuer à la stabilisation de cette société, mais si, dans les faits, les intérêts de certains groupes sont mis à l'écart, un potentiel d'agitation sociale se manifestera alors sous différentes formes (l'aliénation, une baisse du moral, un déclin de l'efficacité de l'autorité, et ainsi de suite) bien avant que des changements se produisent dans les croyances à propos de ce qui est juste [*justness*] relativement à cette société. Cela nous permettra, en outre, d'expliquer la raison pour laquelle les membres de certains groupes en viennent à croire que cette société est injuste, si ceux-ci croient effectivement que c'est le cas.

Outre le fait qu'elles sont caractérisées par un certain type de potentiel d'agitation sociale, les sociétés qui ne parviennent pas à se rapprocher de la rationalité sociale peuvent également partager des caractéristiques supplémentaires : elles peuvent manifester une tendance vers certaines doctrines religieuses ou idéologiques, ou vers certains types d'appareils de répression; elles peuvent être moins

productives à certains égards (par exemple, ne pas être en mesure de développer certaines ressources humaines) et plus productives dans d'autres contextes (par exemple, dans l'extraction à moindre coût d'une plus grande force de travail à partir de certains groupes), et peuvent donc être économiquement prospères de manière différentielle selon les conditions de production avec lesquelles elles doivent composer, etc.

Si une notion de rationalité sociale peut de manière légitime être partie intégrale des explications empiriques des phénomènes de ce type, il doit alors être possible de formuler une caractérisation informative des circonstances dans lesquelles les déviations par rapport à, ou les approximations de la rationalité sociale peuvent être considérées comme conduisant vraisemblablement à des conséquences [*outcomes*] sociales particulières; notamment, une caractérisation informative des conditions dans lesquelles des groupes dont les intérêts sont sacrifiés pourraient vraisemblablement manifester ou mobiliser une forme de mécontentement. Bien qu'on ne puisse pas savoir a priori si une explication de ce type est possible, on peut néanmoins constater dans le cadre des travaux récents en histoire sociale et en sociologie historique l'émergence des divers éléments d'une théorie visant à expliquer quand et comment un potentiel persistant de mécontentement social en vient à se manifester suite au fait que des intérêts persistent à être sacrifiés.³²

Un individu dont les désirs [*wants*] ne reflètent pas ses intérêts ou qui ne parvient pas à la rationalité instrumentale peut, comme je l'ai fait valoir, faire l'expérience d'une rétroaction qui favorise à la fois l'apprentissage en ce qui concerne son propre bien et le développement de stratégies plus rationnelles. De même, le mécontentement suscité par les déviations quant à la rationalité sociale peut produire une rétroaction qui, à l'échelle sociale, favorise le développement de normes qui se rapprochent davantage de la rationalité sociale. Le potentiel d'agitation sociale qui existe lorsque les intérêts d'un groupe sont mis à l'écart est un potentiel de pression pouvant être exercé par ce groupe – et ses alliés – dans le but d'obtenir une reconnaissance plus complète de leurs intérêts dans la prise de décision sociale et dans les normes socialement inculquées, régissant la prise de décision individuelle. Il s'agit donc d'une pression qui oriente davantage la résolution de conflits dans une direction présupposée par la rationalité sociale, puisqu'il s'agit d'une pression qui consiste à donner davantage de poids aux intérêts de celles et ceux qui sont les plus concernés. Une telle pression peut bien évidemment être plus ou moins déterminée ou cohérente. Elle peut se manifester sous la forme d'une expression idéologique bien diversifiée. Elle peut aussi engendrer des conséquences qui sont plus ou moins utiles et favorables à celles et ceux qui exercent cette pression.³³ Des exemples historiques frappants de mobilisation où des groupes exclus ont fait valoir une représentation plus importante de leurs intérêts sont les rébellions contre le système de propriétés féodales, et plus récemment, les mouvements sociaux contre les restrictions à l'égard des pratiques religieuses, du suffrage et d'autres droits civiques, et des négociations collectives.³⁴

Bien évidemment, d'autres mécanismes ont exercé de manière simultanée une influence sur l'évolution des pratiques et des normes sociales, certains desquels ont eu pour conséquence l'effet contraire.³⁵ La question de savoir si les mécanismes qui agissent en faveur de l'inclusion des intérêts exclus prédomineront dépend d'un éventail complexe de facteurs sociaux et historiques. Il serait bête de penser que les normes de n'importe quelle société actuelle à une quelconque étape de l'histoire se rapprochent étroitement de la rationalité sociale, ou bien qu'il y ait une tendance univoque vers un accroissement de rationalité sociale. À l'exemple des mécanismes à l'œuvre dans l'évolution biologique ou dans les économies de marché, les mécanismes décrits ici opèrent dans un « système ouvert » parmi d'autres mécanismes, et rien ne garantit qu'ils conduisent à l'optimalité ou même à un équilibre au sens d'une approche monotone. Les sociétés humaines ne semblent pas avoir été commencées à un état d'équilibre ou à un état qui s'y rapproche, de sorte que la thèse la plus radicale consisterait à affirmer que, à long terme, en l'absence de certains effets exogènes, on pourrait s'attendre à une tendance irrégulièrement séculière vers l'inclusion des intérêts (ou des intérêts représentés par) des groupes sociaux capables d'un certain degré de mobilisation. Mais dans d'autres circonstances, on pourrait s'attendre même à long terme à ce que le contraire se produise. L'esclavage dans le contexte des plantations du Nouveau Monde – assurément, l'une des organisations sociales fondées sur l'exclusion la plus brutale qui n'ait jamais existé – a émergé tardivement dans l'histoire mondiale et a persisté pendant plusieurs siècles. D'autres organisations sociales fondées sur l'exclusion brutale, qu'elles soient anciennes ou récentes, persistent toujours.

Il n'est pas nécessaire, par conséquent, d'épouser une théorie du progrès moral afin de constater que le mécanisme de rétroaction que j'ai décrit jusqu'à maintenant peut accorder un rôle explicatif à la notion de rationalité sociale. L'une des objections les plus déconcertantes, et pourtant la plus courante, adressée au réalisme moral consiste à souligner l'absence de progrès historique uniforme en vue d'un consensus mondial sur les normes morales. Mais l'argument suivant lequel, par exemple, certaines cultures et sous-cultures contemporaines n'acceptent pas – et ne semblent pas être en voie d'accepter – la vision scientifique du monde n'a jamais, à ma connaissance, été invoqué contre le réalisme *scientifique*. Les défenseurs du réalisme scientifique ne sont, bien sûr, pas tenus d'affirmer, dans les deux cas, que des pratiques dans certaines circonstances tendent à produire des théories qui s'appliquent davantage à la réalité; en particulier, lorsque l'objet d'étude est considérablement complexe et si éloigné de la réalité telle qu'elle nous est donnée dans l'inspection directe. Il n'est pas nécessaire qu'ils souscrivent à l'idée désuète voulant que « la vérité éclatera au grand jour » quoiqu'il arrive. La théorie étendue de la rationalité individuelle, par exemple, nous conduit à présager que, dans les sociétés où il existe des conflits d'intérêts considérables, d'une part, les individus développeront des désaccords normatifs importants et, de l'autre, si ces conflits d'intérêts s'établissent en parallèle avec des différences considérables de pouvoir (comme c'est généralement le cas), il est peu probable que les perspectives normatives dominantes soient l'expression de la rationalité sociale. Il est question ici, comme dans le cas de toutes

les explications critérielles en général, d'une explication de certaines tendances [*patterns*] parmi d'autres, et non nécessairement de l'existence d'une seule tendance globale. L'existence des mécanismes de rétroaction, tels que décrits ici, semble toutefois nous offrir des motifs raisonnables de croire que nous pouvons nous rapporter de manière appropriée [*qualified*] à des expériences historiques à titre de quelque chose comme une preuve [*evidence*] expérimentale à propos des types de pratiques dans lesquels un éventail de circonstances pourrait mieux satisfaire à un critère de rationalité sociale. Autrement dit, nous pouvons assigner à ce mécanisme un rôle dans un processus adéquat [*qualified*] d'apprentissage moral.

Les mécanismes d'apprentissage en matière de rationalité individuelle – faible ou étendue – se prêtent à des remarques similaires. Car même si nous nous attendons à ce que, dans des conditions favorables, des individus puissent s'améliorer et mieux agir dans une perspective rationnelle-instrumentale à mesure qu'ils acquièrent plus d'expériences, nous sommes pleinement conscients du fait qu'il existe des mécanismes puissants qui favorisent des résultats opposés. Je ne crois certainement pas qu'un individu doit faire preuve d'une rationalité sans exception, ou même faire preuve d'une rationalité continuellement croissante au cours de sa vie, afin d'appliquer des explications consistant à donner des raisons [*reason-giving explanations*] à plusieurs de ses actions. Je ne pense pas, non plus, que la persistance inévitable de domaines d'irrationalité chez certains individus représente une raison de nier que ceux-ci peuvent, au cours de l'expérience, acquérir des domaines de rationalité plus substantielle.

La comparaison avec la rationalité individuelle ne devrait pas être poussée trop loin. D'abord, bien que les mécanismes qui génèrent l'inclusion dans le cadre de la rationalité sociale agissent par l'intermédiaire de comportements d'individus, des dynamiques interpersonnelles entrent inéluctablement en ligne de compte de telle sorte que les critères sélectionnés ne sont pas réductibles à ceux à l'œuvre dans la rationalité individuelle désagrégée. L'évolution sociale et biologique implique des mécanismes de sélection qui favorisent des comportements satisfaisant aux critères d'optimalité relative, qui sont collectifs (tels que dans le cas du dilemme du prisonnier) ou génotypiques (lesquels peuvent être également collectifs, tel que dans le cadre de la sélection de parentèle), tout aussi bien qu'individuels ou phénotypiques. Si tel n'était pas le cas, il n'aurait guère été possible que des normes morales émergent ou qu'elles aient la prise qu'elles ont désormais sur nous.

Deuxièmement, il existe des différences de degré plutôt extrêmes entre les cas individuels et les cas sociaux. De manière plus frappante, les mécanismes par lesquels les désirs [*wants*] et les comportements individuels sont mis en harmonie avec les intérêts et les raisons individuels agissent de manière plus directe et fiable que les mécanismes analogues qui orientent les pratiques et les normes sociales en direction de ce qui est socialement rationnel. Non seulement les demandes d'information sont moins formidables dans le cas individuel – c'est le moins que l'on puisse affirmer –, mais les façons par lesquelles la rétroaction

est réalisée tendent plus vraisemblablement dans le cas individuel à favoriser le changement et à être moins susceptibles d'être déformées par des asymétries sociales.

Néanmoins, nous avons ici l'esquisse d'une théorie explicative qui emploie la notion de ce qui est plus ou moins rationnel d'un point de vue social et qui, dans le cadre d'explications des croyances et des comportements individuels, reflète très clairement et de manière parallèle le recours à des évaluations de la rationalité du point de vue de l'agent. Tout comme dans la théorie individuelle, elle suggère des généralisations s'appuyant sur des prédictions et des contrefactuels du type suivant : au fil du temps, et dans certaines circonstances plus que dans d'autres, nous devrions nous attendre à ce qu'une pression soit exercée en faveur de pratiques qui satisfont plus adéquatement un critère de rationalité.

Si l'on a ici une théorie potentiellement prédictive et explicative, quelle en est la valeur? Il s'agit là d'une question d'envergure, dont la réponse dépasse le cadre de mes compétences. Mais nous pouvons brièvement noter trois tendances [*patterns*] dans l'évolution des normes morales qui semblent, selon moi, corroborer les prédictions de cette théorie, lesquelles sont soumises aux restrictions liées à l'existence d'imperfections et de mécanismes concurrents. Je les expose toutefois avec une certaine appréhension, car bien que les tendances [*patterns*] que je décris ici soient des tendances historiques grossières, il n'appartient pas nécessairement à la théorie que l'histoire démontre ces tendances, et il ne s'ensuit pas que la théorie doive souscrire à une série de pratiques ou de normes simplement puisqu'elle en résulte.

La généralité. Il s'agit d'un lieu commun en anthropologie que les peuples tribaux ont souvent un terme pour désigner à la fois leur tribu et « le peuple » [*the people*] ou « l'humanité ». Celles et ceux qui ne font pas partie de la tribu ne sont pas considérés comme des individus à part entière, et le type d'obligations que l'on se doit d'avoir auprès des individus ne s'applique pas pleinement aux étrangers. Au cours de l'histoire, par l'entremise de processus qui ont connu de nombreux revirements, les individus se sont accumulés de manière à former des unités sociales plus vastes – de la bande familiale à la tribu, en passant par le « peuple » jusqu'à l'État-nation – et l'étendue des catégories morales s'est élargie afin de suivre cet élargissement des frontières. Il va sans dire qu'il ne s'agit pas là d'une propagation contagieuse des Lumières. Les entités sociales en expansion assujettissent souvent ceux qui se trouvent à l'intérieur de leurs nouvelles frontières, et ceux qui se trouvent dès lors en situation d'oppression parviennent à se garantir une plus grande reconnaissance par des moyens considérablement conflictuels, qui demeurent – et demeureront peut-être toujours – incomplets. Néanmoins, la théorie morale contemporaine et, à un degré qui peut surprendre, le discours moral contemporain en sont venus à rejeter toute forme de limitation au-delà de l'espèce elle-même.³⁶

L'humanisation. Diverses origines et natures ont été assignées aux principes moraux : soit en tant que commandements qui proviennent d'une origine surna-

turelle, fondés dans la volonté ou le caractère d'une déité et devant être interprétés par des prêtres; soit en tant qu'exigences formelles dérivées d'un code d'honneur ancré dans un système de castes; soit en tant que principes d'ordre cosmique; soit en tant que dictats de la raison ou de la conscience, qui ne font appel ni aux inclinations humaines, ni au bien-être, etc. Bien que des vestiges de ces idées survivent dans le cadre de la théorie morale contemporaine, presque toutes les théories morales contemporaines et la majeure partie du discours moral contemporain en général établissent une relation intrinsèque entre les principes normatifs et les effets qu'ils ont sur les intérêts humains. En effet, l'émergence de la moralité en tant que sujet distinct de la religion témoigne de cette tendance.

Les tendances [patterns] de variation. En plus de constater des tendances [patterns] qui reflètent une forme de pression en direction d'une approximation de la rationalité sociale, nous devrions être mesure de remarquer une plus grande approximation de ce genre dans les domaines de régulation normative où les mécanismes postulés ici fonctionnent le mieux, tels que, par exemple, dans les domaines où presque tous les individus ont des intérêts considérablement similaires ou mutuellement dépendants du point de vue de leur satisfaction; où presque tous les individus peuvent éventuellement porter atteinte de manière substantielle aux intérêts d'autrui; où les avantages liés à certaines formes de contrainte ou de coopération sont considérablement saillants même dans le cadre de dynamiques de petits groupes, et où les individus peuvent influencer largement, selon toute vraisemblance, les comportements d'autrui orientés par des normes en obéissant eux-mêmes à des normes. Les exemples les plus évidents ont trait à l'interdiction de l'agression, du vol et des violations de promesses.³⁷ En revanche, il est moins probable qu'une approximation rapide et stable de la rationalité sociale soit réalisée lorsqu'il s'agit de questions morales liées à des situations où il n'existe aucune solution compatible avec la protection des intérêts les plus fondamentaux de tous; où il existe des asymétries très considérables en ce qui concerne la capacité de porter atteinte aux intérêts; où les gains et les pertes qui découlent de certaines formes de coopération ou de contrainte sont difficiles à percevoir; et où la conformité personnelle aura peu d'effet sur la conformité générale. À titre d'exemple, nous pouvons citer des cas qui ont trait à des questions de hiérarchie sociale – par exemple, la permissibilité de l'esclavage, de gouvernements autoritaires, d'inégalités entre les castes et les genres – et de responsabilité sociale – par exemple, quelle est la nature de nos obligations individuelles ou collectives en ce qui concerne la promotion du bien-être des individus avec lesquels nous n'entretenons aucun lien?

Étant donné une caractérisation adéquate des conditions qui ont prévalu au cours des processus d'évolution normative que ces tendances [patterns] décrivent, la présente théorie affirme non seulement que ces changements pouvaient être anticipés, mais qu'une part essentielle de leur explication renvoie à un mécanisme par lequel les individus dont les intérêts ne sont pas pris en considération sont menés à former des valeurs communes et à faire cause commune en vue d'intérêts communs, de sorte qu'ils exercent ainsi une pression sur les pratiques sociales de manière à ce qu'elles en viennent à se rapprocher plus étroitement à la rationalité sociale.

Il ne fait aucun doute que, selon certains, ces descriptions et ces explications de certaines des caractéristiques principales de l'évolution des normes morales semblent, au mieux, naïves, voire, au pire, dangereusement fausses. Cela, je le comprends entièrement. J'ai mis de l'avant des explications douteuses, unilatérales et simplistes à propos d'une réalité très complexe.³⁸ Je ne peux qu'espérer que ces explications semblent aussi crédibles que ce à quoi on peut s'attendre d'explications douteuses, unilatérales et simplistes, et que cela rendrait beaucoup plus plausible l'histoire que j'ai tenté de raconter à propos des mécanismes et des explications.

Il va sans dire que ce qui en résulte n'est pas un fonctionnalisme complaisant ou une approbation globale des pratiques ou des normes morales actuelles. Au contraire, la conception de la moralité esquissée ici met davantage l'accent sur le conflit que sur l'équilibre, et elle nous offre les moyens de critiquer certaines pratiques et intuitions morales contemporaines en posant la question de leur genèse historique. Si nous en venons à penser, par exemple, que l'explication d'une intuition morale commune n'assigne aucun rôle considérable aux mécanismes qui pourraient vraisemblablement exercer une pression en vue de résultats [*outcomes*] socialement rationnels, il s'agit alors d'une raison valable pour remettre en question cette intuition, et ce, malgré le fait que nous y tenons fermement. Dans l'esprit d'une épistémologie morale naturalisée, nous pouvons nous demander si l'explication consistant à élucider la raison pour laquelle nous émettons des jugements moraux constitue un exemple de processus fiable qui nous permet de découvrir des faits moraux.

V. LIMITATIONS

J'ai discuté jusqu'ici de l'excellence morale [*morally best*] sous l'angle de ce qui est rationnel d'un point de vue instrumental et social. Mais j'ai également caractérisé un point de vue de véritablement moral comme ce qui est impartial relativement aux intérêts de tous les individus potentiellement concernés, une notion qui n'est pas socialement conditionnée [*socially-bound*]. En fait, j'ai affirmé qu'une tendance qui va à l'encontre de la spécificité sociale correspond à l'une des tendances [*patterns*] constatables dans le cadre de l'évolution des normes morales. L'explication de ces tendances [*patterns*] – et, par conséquent, le rôle explicatif des degrés de rationalité impartiale – consiste à faire valoir que les mécanismes évoqués ci-dessus ne sont pas également conditionnés sur le plan social [*socially-bound*]. Les sociétés, ainsi que les individus situés aux côtés opposés des frontières sociales, se contraignent mutuellement de diverses manières, tout comme les groupes et les individus se contraignent mutuellement au sein des sociétés : ils peuvent exprimer la menace de commettre une agression, mobiliser une résistance à l'égard des formes de contrôle externe, refuser la coopération et encombrer les plans d'autrui, et ils sont enclins à recourir à ces activités contraignantes lorsque leurs intérêts ne sont pas pris en considération ou sont en danger. Comme pour la moralité intrasociale, les éléments rationnels les mieux établis et presque les plus impartiaux dans le cadre de moralité inter-sociale sont ceux qui sont à l'œuvre lorsque les mécanismes évoqués fonction-

nent de manière optimale [*most reliably*] : les interdictions d'agression sont plus contraignantes et plus largement acceptées que les principes d'équité ou de redistribution. Bien entendu, il y a plusieurs facteurs qui font en sorte que les dynamiques intersociétales diffèrent des dynamiques intrasociétales... Mais j'épargnerai pour une fois le lectorat de toute forme de science sociale de comptoir. Et pourtant, ce qui en ressort est une forme de réalisme moral qui est essentiellement lié à un point de vue limité, à un point vu impartial et pourtant humain. S'agit-il d'une perspective trop limitée pour avoir un réalisme moral authentique?

Un de mes professeurs a déjà souligné que la question du réalisme moral semblait, selon lui, se ramener à la question de savoir si ce que nous faisons importe ou non à l'univers. Puisque nous avons renoncé depuis bien longtemps à croire en l'idée que nous ne sommes pas indifférents au cosmos, il pensait que nous avions par le fait même renoncé au réalisme moral. Je ne peux être d'accord avec ce jugement que si le réalisme moral implique en effet cette idée; on doit alors – avec soulagement plutôt qu'avec tristesse – y renoncer. Toutefois, l'explication proposée ici nous offre un moyen de comprendre comment les valeurs ou les impératifs moraux peuvent être objectifs sans être cosmiques. Ils n'ont pas à être fondés dans rien de plus transcendantal que des faits à propos de l'être humain et de son environnement, des faits concernant les sortes de choses qui nous importent, et en quoi nos façons de vivre ont une incidence sur ces choses.

Or la présente théorie implique une limitation en un tout autre sens, laquelle peut être une source d'inquiétude beaucoup plus importante du point de vue de la théorie morale contemporaine : elle ne donne pas lieu à des impératifs moraux catégoriques au sens où elle nous fournirait une raison d'agir pour tous les agents rationnels indépendamment de leurs désirs contingents. Bien qu'elle soit troublante, cette limitation ne conduit pas au relativisme étant donné que, selon la présente théorie, la motivation rationnelle n'est pas une condition préalable de l'obligation morale. Par exemple, il pourrait être affirmé honnêtement que nous devrions être plus généreux même si une plus grande générosité ne nous aidait pas à promouvoir nos fins existantes ou même à satisfaire nos intérêts objectifs. Il pourrait en être ainsi parce que l'action moralement bien [*right*] que nous devons poser dépend de ce qui est rationnel d'un point de vue qui implique le nôtre, mais sans toutefois se limiter à lui.

De manière similaire, on pourrait affirmer que nous ne devrions pas croire à la fois à une proposition *p* et à une proposition qui implique non-*p*. Cependant, il se peut que chaque agent rationnel n'ait aucune raison instrumentale de purger toutes les contradictions logiques de son esprit. Une telle tâche exigerait pour quiconque une part considérable de réflexion si chacun devait tester la cohérence de toutes ses croyances existantes et s'assurer que chacune de ses croyances nouvellement acquises respecte cette cohérence. Supposons que quelqu'un soit si privilégié que les seules contradictions dans ses croyances résident en profondeur dans le marécage considérablement sédimenté de ses

connaissances factuelles d'ordre général. Il se peut que ses souvenirs au sujet de deux connaissances [*acquaintances*] de son passé soient devenus si entremêlés les uns avec les autres que quelque part dans la boue se trouvent deux croyances distinctes qui, mises ensemble, attribuent à un même individu des propriétés logiquement incompatibles. Jusqu'au moment où une telle contradiction surgit de manière explicite en pratique, l'individu ne peut avoir aucune raison d'explicitement ses préoccupations actuelles et d'y remédier après coup, pas plus qu'il n'a de raison de quitter son domicile dans la banlieue du New Jersey pour aller chasser des alligators dans le marais d'Okefenokee, au cas où il pourrait se trouver un jour perdu et sans arme dans les remous de la Géorgie du Sud-Est.³⁹ Ce qu'un individu devrait rationnellement faire peut ainsi différer de ce qui s'impose logiquement à lui. Néanmoins, nous pouvons affirmer que les évaluations logiques ne sont pas subjectives ou arbitraires et qu'il y a de bonnes raisons, parfaitement générales, pour qu'un individu soit logique, notamment que les contradictions logiques sont nécessairement fausses et les inférences logiques sont un gage de vérité. Étant donné que dans le cadre du discours public et dans nos réflexions personnelles nous sommes souvent concernés par la question de savoir si notre pensée est justifiée en un sens qui est lié davantage à ses conditions de vérité qu'à sa valeur instrumentale relative à nos buts personnels, le fait que nous attachions autant d'importance à la logique à titre de standard de critique et de critique de soi est donc loin d'être arbitraire.

Parallèlement, si nous adoptons la conception du bien moral [*moral rightness*] proposée ci-dessus nous pouvons affirmer que les évaluations morales ne sont pas subjectives ou arbitraires, et qu'il existe de bonnes raisons générales pour respecter des obligations [*oughts*] morales, notamment que la conduite morale est rationnelle d'un point de vue impartial. Étant donné que dans le cadre du discours public et dans nos réflexions personnelles nous sommes souvent concernés par la question de savoir si notre conduite est justifiable d'un point de vue général plutôt que d'un point de vue simplement personnel, le fait que nous attachions autant d'importance à la moralité à titre de standard de critique et de critique de soi est donc loin d'être arbitraire.

L'existence de phénomènes tels que la religion et l'idéologie attestent de l'omniprésence et de l'importance de notre souci pour la justification impartiale. Au cours de l'histoire, des individus ont sacrifié leurs intérêts et même leurs vies afin de satisfaire aux exigences des religions et des idéologies qui étaient contraignantes pour eux en partie puisqu'elles prétendaient exprimer un point de vue universel – l'universel – d'un point de vue justificatif. La Rochefoucauld écrit que l'hypocrisie est un hommage que le vice rend à la vertu,⁴⁰ mais « hypocrisie » connote ici le cynisme. Il serait plus approprié d'affirmer que l'idéologie est le respect que les partisans accordent à l'impartialité. La moralité n'est pas alors l'idéologie devenue sincère et générale – une idéologie est intrinsèquement portée à des généralisations sincères [*given to heart-felt generalization*]. La moralité est l'idéologie qui a fait face aux faits.

Nous soupçonnons que l'idée selon laquelle les évaluations morales doivent posséder une force catégorique pour des agents rationnels est en partie attribuable à la crainte que l'autorité de la moralité disparaisse si cette idée était niée. Cela serait effectivement le cas si on tenait à l'idée que les impératifs moraux ne pourraient exister pour quelqu'un s'il n'avait aucune raison d'y obéir, car un individu pourrait alors échapper de manière expéditive aux devoirs moraux par le fait d'avoir des désirs bas. Mais si nous abandonnons cette idée concernant l'applicabilité des jugements moraux, des variations dans les désirs personnels ne peuvent alors cautionner aucune exemption sur le plan des obligations morales.⁴¹

Par conséquent, bien que l'argument présenté ici ne parvienne pas à formuler une conception des impératifs moraux en tant qu'impératifs catégoriques, il s'agit là d'une limitation avec laquelle nous pouvons composer, et qui peut néanmoins consacrer à la moralité l'étendue et la dignité dont elle jouit traditionnellement. Car combien parmi nous peuvent parvenir à se convaincre que la raison ne peut être autre qu'hypothétique? Doit-on également poser la question de savoir combien d'individus parmi nous considéreraient que l'importance que nous accordons à la moralité ou à la conduite morale serait augmentée par une démonstration suivant laquelle même une personne ayant les fins les plus répugnantes trouvera que la conduite morale parvient à les promouvoir?

Une des implications de la théorie présentée ici est que si nous voulons que la moralité soit prise au sérieux et qu'elle ait une place importante dans la vie des individus – et non pas simplement en tant qu'illusion ou sous l'effet d'une menace de répression – nous devrions alors nous soucier considérablement des diverses modalités par lesquelles les arrangements sociaux produisent des conflits d'intérêts et des asymétries de pouvoir, qui ont une incidence sur la nature et la taille de l'écart qui sépare la rationalité individuelle et la rationalité sociale. Au lieu de tenter de concevoir la moralité comme ce qu'elle ne peut être, c'est-à-dire comme « ce qui est rationnellement contraignant peu importe les fins de tout et chacun », nous devrions plutôt nous demander comment nous pouvons changer nos modes de vie de manière à ce que la conduite morale soit plus régulièrement rationnelle étant donné les fins que nous avons actuellement.

VI. RÉSUMÉ ET CONCLUSION

J'ai esquissé une forme de réalisme moral et indiqué comment elle peut être défendue contre certaines objections. Je ne peux présenter dans le cadre du présent essai ni une caractérisation complète de cette théorie, ni des réponses complètes aux nombreuses objections qu'elle suscite. Je devrais peut-être alors cesser de me prononcer sur ces questions et clore cet essai en indiquant approximativement ce que j'ai tenté de présenter et ce que je n'ai pas exposé.

J'ai proposé de réformer les définitions naturalistes de ce qui est non moralement bon [*goodness*] et de ce qui est moralement bien [*moral rightness*]. On pourrait me répondre : « Oui, nous pouvons constater qu'une telle ou telle fin est un inté-

rêt objectif de l'agent au sens où vous l'entendez, ou qu'une telle ou telle pratique est rationnelle d'un point de vue impartial, mais pouvons-nous néanmoins nous demander si la fin est bien pour elle ou lui ou si la pratique est bonne [*right*]? » Or de telles « questions ouvertes » ne peuvent de par leur nature être résolues, étant donné que les définitions ne font pas l'objet de démonstrations ou de réfutations. Mais de telles questions ouvertes peuvent être plus ou moins troublantes, car bien que des propositions définitoires ne puissent être démontrées, elles peuvent mieux s'en tirer ou non à la lumière de divers *desiderata*.

J'ai présupposé que le travail d'élaboration de définitions consiste en une partie de la construction théorique et se doit d'être évalué en se demandant 1) si les analyses satisfont aux contraintes adéquates d'intelligibilité et de fonctionnalité, et 2) si les termes en tant qu'ils sont analysés contribuent à la formulation et à la vérification des théories recevables. En quoi notre théorie répond-elle à ces critères?

1) Outre les contraintes d'intelligibilité, telles que la clarté et la non-circularité, les définitions spécifiquement naturalistes de termes évaluatifs doivent satisfaire deux autres contraintes qui relèvent de leurs fonctions spécifiques. a) Elles doivent dans la mesure du possible saisir la force normative de ces termes en proposant des analyses qui permettent à ces termes de jouer leurs rôles évaluatifs fondamentaux. Dans le cadre actuel, cela signifie de démontrer qu'elles expriment néanmoins des notions de ce qui est bon [*goodness*] et de ce qui est bien [*rightness*] que l'on peut reconnaître, bien que les définitions proposées ne correspondent pas à toutes nos intuitions linguistiques et morales. Par ailleurs, cela signifie de démontrer que les définitions permettent d'établir des liens plausibles entre, d'une part, ce qui est bon ou bien et, de l'autre, ce qui typiquement motiverait les individus prêts à se soumettre à des formes adéquates d'examen approfondi. b) Les définitions naturalistes doivent s'assurer que les concepts évaluatifs participent, de plein droit, à des théories véritablement empiriques. Il s'agit, d'une part, de montrer que nous avons un accès adéquatement épistémique à ces concepts. Il s'agit, d'autre part (et corrélativement), de montrer que les généralisations recourant à ces concepts, entre autres, peuvent intervenir dans le cadre de théories potentiellement explicatives. J'ai tenté de proposer des définitions raisonnablement claires et de montrer à titre préliminaire comment elles peuvent satisfaire les contraintes a) et b).

2) Il reste toutefois encore beaucoup à faire, car il faut encore démontrer que les théories empiriques élaborées à l'aide de ces définitions sont des théories raisonnablement bonnes, c'est-à-dire des théories qui s'appuient sur des preuves substantielles et qui donnent lieu à des explications plausibles. C'est ce que j'ai tenté de suggérer à titre d'étape strictement préliminaire. Si j'ai été tout à fait peu convaincant sur les sujets d'ordre empirique, je m'attends alors à ce que les définitions proposées ici soient aussi peu convaincantes.

Un atout du naturalisme en éthique et en épistémologie tient, selon moi, au fait qu'il présente des contraintes dans plusieurs domaines importants à la fois.

Plusieurs occasions s'offrent à nous pour démontrer que la théorie est fautive ou est jugée peu convaincante lorsqu'elle doit répondre à des exigences empiriques ainsi qu'à des intuitions normatives. L'effort de théorisation en général s'avère plus productif lorsque des contraintes adéquates s'imposent; en particulier, des contraintes sont exigées en éthique si nous devons avoir une idée plus claire quant à la façon de faire des progrès en ce qui concerne la résolution théorique de conflits. Bien sûr, il ne s'agit pas simplement de n'importe quelle contrainte. L'ensemble des contraintes proposées doit se présenter à la fois comme adéquat et utile. Pour conclure, nous ferons quelques remarques au sujet de 1) l'utilité des contraintes adoptées ici et 2) sur leur pertinence.

1) Considérons trois classes de théories qui concurrencent la théorie morale substantielle qui a été exposée ci-dessus, et remarquons comment les critiques qu'on leur adresse confondent naturellement des préoccupations relatives à la justification normative et à l'explication empirique. On considère généralement que les conceptions *kantiennes* de la moralité rendent compte de certaines caractéristiques normatives, intuitivement probantes au sujet de notions telles que la rationalité et le bien moral [*moral rightness*], mais il semble que c'est possible, en partie, par l'omission de l'élaboration d'un moyen plausible par lequel ces notions seraient intégrées dans une théorie empirique des raisons et des motifs d'agir. Qui plus est, cette difficulté descriptive trouve son expression directe dans le domaine normatif. Non seulement une « obligation » [*ought*] normative doit relever d'une « capacité » [*can*] empirique, mais une obligation normativement convaincante [*compelling*] doit – comme l'ont souligné les critiques récentes du kantisme – atteindre les véritables sources de l'action et des préoccupations humaines. Les théories morales *intuitionnistes* ont également joui d'un certain succès par rapport au fait de saisir certaines caractéristiques normatives de la moralité, mais elles ont considérablement été abandonnées faute de fournir une théorie crédible de la nature et de l'opération de la faculté de l'intuition morale. Il nous est trop facile de formuler une explication psychologique non justificative de l'existence en de quelconques gentlemen anglais de ce que ceux-ci parviennent à identifier au moyen de l'introspection comme une faculté de l'intuition morale, une explication qui lie cette prétendue faculté plus étroitement avec la rigidité des conventions sociales existantes qu'avec quelque chose qui pourrait correspondre en apparence à une source de vérité universelle.

2) L'enjeu qui se pose toutefois lorsque nous évaluons des approches concurrentes de la moralité implique non seulement une forme d'évaluation des théories importantes, mais également des questions concernant le type de critères d'évaluation qui s'appliquent aux définitions et aux théories en éthique, de même que le fait de déterminer si l'effort considérable de systématisation définitionnelle et de théorisation est tout simplement approprié en ce qui concerne l'éthique. Je suis porté vers l'idée selon laquelle le développement des théories en éthique n'est pas une construction artificielle des philosophes, mais le résultat organique d'usages personnels et sociaux de l'évaluation morale : les individus et les groupes ont fait face, maintes et maintes fois, à des questions difficiles auxquelles le sens commun ne pouvait que donner des réponses

conflictuelles ou sinon insatisfaisantes, de sorte qu'ils ont approfondi davantage leurs questions et ont poursuivi leur enquête de manière plus systématique. Ils ont ressenti le besoin de formuler une théorie en éthique qui s'articulerait en parallèle avec le besoin éprouvé d'élaborer une théorie relevant des sciences naturelles ou sociales.⁴² Il n'en résulte pas que seule la théorisation éthique doit s'articuler en parallèle avec – ou s'intégrer dans – la théorisation dans le cadre des sciences naturelles et sociales. Bien qu'elle puisse sembler initialement plausible et ultimement irréfutable, l'idée suivant laquelle l'éthique doit être isolée est également une idée que nous rejetons ultimement. Nous sommes des êtres naturels et sociaux, et nous n'avons aucune idée où nous pourrions regarder ailleurs, en ce qui concerne l'éthique, que dans cette riche conjonction de faits. J'ai tenté de suggérer que c'est dans ces domaines que nous pourrions effectivement trouver une éthique.⁴³

REMERCIEMENTS

Traduction de Peter Railton, “Moral Realism”, *The Philosophical Review*, Vol. 95, No. 2, p.163-207, par Denis Courville. *Les ateliers de l'éthique/The Ethics Forum* souhaite remercier *Duke University Press* de nous avoir accordé la permission de traduire et de publier ce texte ainsi que Denis Courville pour en avoir fait la traduction.

NOTES

- ¹ Le naturalisme évolutionnaire du XIX^e siècle offre un exemple de la première variante, alors que Dewey – et, du moins selon une certaine interprétation, peut-être Mill également – offre un exemple de la seconde.
- ² A. J. Ayer, « Langage, vérité et logique », Paris, Flammarion, 1956, p. 156-157.
- ³ En faisant valoir cette idée, je soutiens que les questions portant sur ce qui est rationnel de croire appartiennent à la raison pratique plutôt qu'à la raison théorique. Bien que les résultats de la raison théorique – par exemple, les conclusions d'inférences déductives – soient en général pertinents pour les questions à propos des croyances rationnelles, ils ne sont pas déterminants, mis à part les raisons pratiques d'un agent.
- ⁴ Bien entendu, la formation de croyances individuelle n'est pas en général gouvernée explicitement par un raisonnement de type moyen-fin, mais plutôt par des habitudes liées à la formation de croyances et par des tendances à investir des degrés variables de confiance dans certaines croyances. Si nous acceptons une conception instrumentale de la rationalité, nous pouvons alors qualifier de rationnelles ces habitudes du point de vue de l'individu, dans la mesure où elles s'insèrent dans une constellation d'attitudes et de tendances qui font la promotion de ses fins. Cette question sera abordée de nouveau dans la section IV.
- ⁵ Ni ces remarques, ni celles formulées dans les paragraphes suivants se veulent une exégèse sérieuse des arguments de Hume, lesquels admettent des interprétations autres que ceux qui sont offerts ici. Je désire seulement rendre compte de certaines caractéristiques de ce que je considère être les arguments de Hume, par exemple, dans le livre III, partie I, section I du *Traité de la nature humaine (La morale)*, Paris, Flammarion, 1993, en particulier p. 60-61, et dans l'annexe I de l'*Enquête sur les principes de la morale*, Paris, Flammarion, en particulier p. 214-215.
- ⁶ Philippa Foot a remis en cause cette thèse, et bien que la façon dont elle pose et fait valoir la question diffère suffisamment de la mienne que je ne puis juger si elle est d'accord avec l'argument qui suit. Voir son ouvrage *Virtues and Vices*, Berkeley, University of California Press, 1978; en particulier l'essai XI. La présentation de cette question s'inspire principalement de la distinction établie par William K. Frankena entre le point de vue rationnel et le point de vue moral.
- ⁷ Voir Hume, David, *Enquête sur les principes de la morale*, section IX, partie II, p. 196.
- ⁸ La critique ancienne du non-cognitivism, selon laquelle il éprouve de la difficulté à rendre compte de la différence entre les valeurs morales et les autres types de désirabilité (de sorte que Hume peut parler dans un même souffle de « bons offices » et de son logis bien disposé) connaît un certain regain de vitalité dans le présent contexte. Il est nécessaire pour rendre compte de différences de ce genre que l'on dispose d'un moyen substantiel [*contentful*] de caractériser des critères d'évaluation morale afin que l'approbation morale ne soit pas réduite à « ce à quoi un agent accorde de l'importance ». (Une caractérisation de ce type sera présentée à la section IV.) Une valeur *sans phrase* [en français dans le texte] est une notion générique, et non nécessairement morale. Il est parfois affirmé que la valeur générique présente un caractère moral lorsque l'on atteint ce à quoi un agent accorde par-dessus tout de l'importance. Mais cela aurait pour effet d'investir de valeur morale les animaux de compagnie et les mémén-tos, et aurait l'effet particulier de rendre presque impossible l'amoralisme. Il semble plus plau-

sible d'affirmer que les valeurs ne sont pas toutes morales, et que les valeurs les plus élevées ne doivent pas nécessairement être – ou même simplement être considérées par un agent comme – des valeurs morales. Lorsque l'on se tourne vers des questions d'obligation [*duty*], la situation devrait être encore plus claire : les théoriciens moraux ont proposé des relations tout à fait différentes entre les catégories à propos de ce qui est moralement bien [*moral rightness*], moralement bon [*moral goodness*], et non moralement bon [*non-moral goodness*], et il ne semble pas plausible d'affirmer que le fait de considérer une action ou une classe d'actions comme moralement bonne [*right*] est nécessairement équivalent au fait de la considérer personnellement comme ce qui a de valeur [*valuable*] *sans phrase*.

⁹ La nature des impératifs moraux est discutée plus longuement à la section V.

¹⁰ Des théories de ce type sont dûment externalistes lorsque, en définissant les notions de *fiabilité* [*reliability*] ou de *processus causaux qui confèrent la justification* [*warrant-conferring causal process*], elles ont recours à une théorie de la vérité qui ne réduit pas celle-ci à ce en vertu de quoi nous avons des raisons de croire – par exemple, une théorie correspondantiste non triviale.

¹¹ Ou si les faits moraux étaient censés d'être le type de choses qui fournit des raisons catégoriques d'agir. Toutefois, cette supposition consiste simplement à revêtir la thèse humienne de la praticabilité d'une tenue ontologique.

¹² J. L. Mackie, in *Ethics: Inventing Right and Wrong*, Harmondsworth, Middlesex, Penguin, 1977; et Gilbert Harman, in *The Nature of Morality: An Introduction to Ethics*, New York, Oxford University Press, 1977. Tous les deux contestent le réalisme moral en partie en remettant en cause sa capacité d'explication. Nicholas L. Sturgeon, in « Moral Explanations », David Copp and David Zimmerman (eds.), *Morality, Reason and Truth: New Essays in the Foundations of Ethics*, Totowa, N.J.: Rowman and Allanhead, 1984, adoptent l'opinion contraire, en ayant recours à des arguments distincts de ceux présentés ci-dessous.

¹³ Je crois qu'une théorie exhaustive de la valeur pourrait démontrer que le concept du bien [*good*] de quelqu'un est légèrement différent du concept de ce qui est désirable pour lui. Cette différence n'aura toutefois aucune incidence sur l'argument présenté ici.

¹⁴ C'était le travail de Richard C. Jeffrey sur les probabilités épistémiques qui m'ont conduit à la base à l'idée d'intérêts subjectifs objectifiés. Voir la note 17. J'ai tiré profit depuis lors du travail de Richard B. Brandt sur les « désirs rationnels », bien que je craigne qu'il considère une part notable de choses affirmées ici comme erronée. Voir la partie I de *A Theory of the Good and the Right*, Oxford, Clarendon, 1979.

¹⁵ Nous posons cette question à A^+ , plutôt que celle de savoir ce que A^+ voudrait pour lui-même, puisque nous sommes à la recherche des intérêts subjectifs objectifiés de A , et les intérêts de A^+ pourrait être tout à fait différents en raison des changements impliqués dans l'idéalisation de A . Par exemple, on peut supposer vraisemblablement que A^+ ne désire pas davantage d'informations au sujet de lui-même, car il n'y a rien de plus à ajouter, et cela il le sait. Il pourrait néanmoins être tout à fait possible que A^+ désirera davantage d'informations s'il était à la place de la version moins informée de lui-même, à savoir A . Cela pourrait être impossible, en tant que question psychologique, pour A^+ de mettre complètement de côté ses désirs dans ses circonstances actuelles au sujet de lui-même ou A , lorsqu'il s'agit de considérer ce qu'il voudrait vouloir s'il se trouvait à la place de la version moins qu'idéale de lui-même. Tout cela révèle un problème de mesure concernant les intérêts objectifs : le fait de mettre à la disposition d'un individu les informations et les capacités nécessaires à l'« objectivation » de ses intérêts peut perturber sa psychologie de manière à ce que le phénomène que l'on désire observer s'en trouve modifié. De telles difficultés se présentent même au cours de la mesure des intérêts subjectifs étant donné que les instruments utilisés pour échantillonner les préférences (en effet, de simples actes de réflexion d'un individu sur ses propres préférences) tendent à avoir une incidence sur les préférences exprimées. Pour des raisons évidentes, les effets d'interférence sont inévitables. Bien qu'ils ne constituent pas en eux-mêmes une raison suffisante pour justifier une forme de scepticisme à l'égard des intérêts subjectifs et objectifs, ces problèmes de mesure illustrent bien l'importance d'une « théorie des perturbations » [*perturbation theory*], et la nécessité de faire preuve de prudence à l'égard des attributions d'intérêts qui se révèlent inattentives aux effets d'interférence.

¹⁶ « Intérêt » n'est pas tout à fait le mot recherché ici, car dans le langage courant nous pouvons parler d'un désir (*want*) là où nous ne pourrions pas parler d'un intérêt correspondant. Cf. Brian Barry, *Political Argument*, Londres, Routledge et Kegan Paul, 1965, en particulier, le chapitre X, pour une discussion de ce genre. Une expression plus précise, mais excessivement encombrante, serait « caractéristique donnant lieu à une valeur positive » [*positive-valence-making characteristic*].

¹⁷ Supposons un instant, contrairement à ce qui a été établi ci-dessus, que nous disposons d'une notion acceptable de probabilité épistémique qui détermine des degrés rationnels de croyance indépendamment des objectifs contingents de l'agent épistémique. Peut-être que l'analogie suivante sera alors utile. Prenons un processus physiquement contingent, comme la désintégration alpha. Nous pouvons demander à un individu quelle probabilité subjective il assignerait à un événement consistant dans une certaine vitesse [*rate*] de dégradation pour un quelconque échantillon d'uranium; nous pourrions aussi demander quel est le degré rationnel de croyance qu'un individu pourrait assigner à cet événement s'il devait alors être idéalement informé des lois de la physique et des conditions initiales pertinentes. Nommons ce degré rationnel de croyance la probabilité subjective objectivée de l'événement, et supposons qu'elle soit égale à un cinquième (à comparer avec : Jeffrey, Richard C., *The Logic of Decision*, New York, McGraw-Hill, 1964, p. 190-196). Mais considérons à présent les faits physiques qui, en lien avec les lois de la mécanique quantique, fondent le jugement idéalisé de l'individu. Appelons-les la base réductive de ce jugement. Cette base réductive consiste dans un ensemble complexe de qualités premières, dont on peut dire qu'elle rend possible le fait qu'à l'événement en question correspond une probabilité objective d'un cinquième (il faut préciser que l'approche de Jeffrey ne postule pas de telles probabilités objectives.) L'existence de cette probabilité objective peut expliquer la raison pour laquelle un individu idéalement informé pourrait sélectionner une probabilité égale à un cinquième, mais le jugement de probabilité d'un individu idéalement informé ne peut expliquer pourquoi la probabilité objective est d'un cinquième, car cela relève des lois de la physique. De même, l'existence d'un intérêt objectif d'un individu peut expliquer la raison pour laquelle la version idéalement informée de cet individu pourrait sélectionner un intérêt subjectif objectifié pour la version moins informée de lui-même, mais non l'inverse.

¹⁸ Plus précisément, nous pourrions affirmer que *X* est non moralement bon [*good*] pour *A* à un temps *t* si et si seulement *X* satisfait un intérêt objectif de *A* lorsqu'une base réductive existe à *t*. Un bon nombre de questions peuvent être soulevées au sujet des considérations relatives à l'évolution des intérêts dans le temps, mais on ne peut les aborder ici.

¹⁹ *A*+, en se mettant à la place de *A*, peut trouver plusieurs ensembles de désirs [*wants*] différents tout aussi attrayants les uns que les autres, de telle sorte que plusieurs options peuvent être, en ce sens, équivalentes-meilleures [*equal-best*] pour *A*. Sans pour autant devenir problématique, la notion du « bien pour *A* » deviendrait alors plutôt pluraliste [*pluralistic*]. Toutefois, une question plus sérieuse se pose : la spécification des conditions de *A*+ ou de la psychologie du désir sont-elles assez déterminées pour que la notion d'intérêt objectif soit suffisamment définie afin de répondre à mes objectifs? Sans préciser quel pourrait être ce niveau de caractérisation, je propose deux façons d'esquisser une réponse au problème de la précision [*definiteness*]. 1) Il semble que nous pensons effectivement qu'il existe des réponses plus précises quant à la façon dont les désirs d'un individu *A* pourraient être modifiés si ses croyances devaient également se modifier sous certaines conditions. Si Lonnie devait apprendre ce qui résulte du fait de boire du lait, il ne voudrait plus que son désir de boire du lait soit opérant. Mais un changement considérable dans les croyances peut s'effectuer petit à petit sur la base d'une séquence de changements limités dans les croyances. Ainsi, si (et il s'agit là, certes, d'un grand « si ») l'ordre du changement n'a aucune importance au final, alors les faits et les généralisations sur lesquels s'appuient les contrefactuels à propos des changements limités pourraient appuyer une extrapolation jusqu'à *A*+. 2) Les croyances et les désirs semblent être sujets à une covariation systématique. En général, nous constatons que des individus qui diffèrent notablement dans leurs désirs – par exemple, concernant leurs carrières ou leur style

de vie – diffèrent notablement, et typiquement, dans leurs croyances; au fur et à mesure que les individus se ressemblent dans leurs croyances, ils tendent à se ressembler dans leurs désirs. Cela laisse à penser que si (un autre grand « si ») la caractérisation proposée de A^+ fixe de part en part le contenu de ses croyances de manière définitive (du moins, étant donné un choix de langage), alors ses désirs peuvent également être complètement fixés. Si nous disposions d'une théorie générale de la covariation des croyances et des désirs, nous pourrions alors faire appel directement à cette théorie – en plus des faits à propos de A – pour déterminer les contrefactuels nécessaires pour caractériser les intérêts objectifiés de A , de sorte que nous pourrions éliminer par là même toute référence essentielle à l'individu imaginaire A^+ .

²⁰ L'explication en question peut, toutefois, donner lieu à des résultats contre-intuitifs. Suivant la nature et les circonstances d'individus donnés, les intérêts objectifs de ceux-ci peuvent consister dans des choses que nous considérons comme mal et répugnantes, et qui ne semblent pas, selon nous, faire partie de ce qu'est une bonne vie. Nous pouvons expliquer une bonne part de nos objections à certains désirs – par exemple, ceux qui impliquent la cruauté – en affirmant qu'ils ne sont pas *moralement* bons [*good*]; d'autres désirs – par exemple, de nature philistine – en affirmant que, d'un point de vue *esthétique*, ils n'ont aucune valeur, et ainsi de suite. Il me semble préférable d'exprimer notre dégoût pour certaines fins en termes de catégories spécifiques de valeur plutôt que d'avoir recours à l'idée suivant laquelle de telles fins ne pourraient en aucune circonstance faire partie du bien non moral [*non-moral good*] de quelqu'un. L'ensemble des personnes – ou du moins certains individus – pourrait être organisé de manière à ce que certaines choses qui ne semblent pas très attrayantes deviennent essentielles à son épanouissement [*flourishing*], et nous ne voulons pas être coupables de vœux pieux à ce sujet (il y aura suffisamment de vœux pieux d'ici la fin).

²¹ Pour reformuler la question en des termes plus strictement naturalistes, nous pourrions nous attendre à ce que l'évolution ait favorisé des organismes constitués de telle sorte que les comportements nécessaires à leur survie et à leur développement [*flourishing*] soient associés à des états internes positifs (tels que le plaisir), alors que les comportements qui s'opposent à leur survie et à leur développement [*flourishing*] soient associés à des états négatifs (tels que la douleur). La notion de « développement » [*flourishing*], même si elle est comprise comme une simple aptitude à la survie [*fitness*] reproductive, n'est pas une notion étroite. Pour que des êtres tels que les êtres humains puissent être reproductivement féconds [*successful*], ils doivent, en tant que phénotypes, avoir des vies qui sont psychologiquement viables [*sustainable*], motivées de manière interne [*internally motivating*], et effectivement sociales. Il s'agirait également de vies qui, en général, se livreraient à un grand nombre de capacités spécifiquement humaines. L'espèce humaine ne pourrait difficilement avoir connu le succès qu'elle a connu, même sur le plan reproductif, si ce n'était pas de la recherche [*pursuit*] des types de choses qui, de manière caractéristique, ont incité les êtres humains à agir en fonction de ce type d'existence. Toutefois, il faut garder à l'esprit que la majeure partie du processus de l'évolution humaine a eu lieu dans des circonstances qui se distinguaient, en de multiples façons, de celles d'aujourd'hui. Il est donc tout à fait possible que l'interaction entre le potentiel motivationnel de l'humanité ait évolué et que les circonstances actuelles donneront lieu, à leur tour, à des incongruités entre ce vers quoi nous tendons, ou ce qui nous anime, et ce qui pourrait produire le plus grand plaisir pour nous tous. Il s'agit là d'une raison pour remettre en doute l'hédonisme en tant que théorie de la motivation.

²² Il est question ici de « succès fonctionnel » plutôt que de « précision représentationnelle » [*representational accuracy*] pour les raisons suivantes. La sélection favorise les organismes qui ont telle ou telle autre caractéristique, qui contribue à répondre à leurs besoins dans le cadre d'environnements particuliers. La question de savoir si à cette caractéristique correspond une capacité représentationnelle précise [*accurate*] ne peut être résolue par un argument de ce type. Ce serait, bien sûr, une très grande coïncidence si des êtres qui se fient aussi considérablement que nous sur des représentations étaient seulement aptes à construire des représentations grossièrement inexactes, tout en étant capables de composer fructueusement [*successfully*] avec divers environnements sur l'étendue d'une longue période de temps. Mais de telles coïncidences ne peuvent être écartées.

²³ Bien que certains éléments de leur base réductive dépendent de nos choix antérieurs, nos intérêts objectifs ne sont pas, par conséquent, subjectifs en un sens qui est préjudiciable à l'argument présenté ici. Après tout, des faits objectifs qui nous concernent et qui sont aussi non problématiques que notre poids, notre revenu et notre position spatiale dépendent de nos choix antérieurs. Il ne s'agit pas ici d'affirmer que nos intérêts subjectifs n'ont aucun rôle dans la constitution de la base réductive de nos intérêts objectifs, mais plutôt que ceux-ci ne peuvent influencer nos intérêts objectifs qu'en vertu des effets actuels (plutôt que des effets simplement désirés) qu'ils ont sur cette base réductive, tout comme nos choix antérieurs ne peuvent influencer notre poids, notre revenu ou notre position spatiale qu'en vertu des effets actuels (plutôt que des effets simplement désirs) qu'ils ont sur notre déplacement, notre situation d'emploi ou nos mouvements.

²⁴ De même, une explication serait incomplète et superficielle si elle tentait d'élucider les raisons pour lesquelles la consommation mondiale de sucre raffiné a connu une croissance fulgurante lorsque la production à grande échelle est devenue possible, en se limitant tout simplement au constat que les gens apprécient le goût du sucre. Pourquoi, malgré la présence d'une forte divergence dans les alimentations traditionnelles et le développement des goûts, le sucre est-il parvenu à faire de si grandes percées dans la consommation humaine? Pourquoi l'apparition et la promotion de produits alimentaires tout aussi peu coûteux n'ont-elles pas provoqué des changements de consommation aussi remarquables? Pourquoi la consommation de sucres, souvent dissimulés sous d'autres formes, continue à croître, et ce, même dans les sociétés où le sucre est reconnu comme présentant un risque pour la santé? Les faits concernant la manière dont nous sommes constitués, les divers moyens singuliers par lesquels le sucre nous affecte, et les façons dont les formes de production et les types de consommation ont conjointement évolué ensemble de manière à générer une demande en croissance et une offre en expansion, doivent suppléer une explication qui s'en tient seulement au niveau des préférences subjectives. Pour une discussion approfondie de cette question, cf. Sidney W. Mintz, *Sweetness and Power: The Place of Sugar in Modern History*, New York, Viking, 1985.

²⁵ Pour une explication de la façon dont les théories informellement exprimées peuvent néanmoins transmettre de l'information explicative, cf. la section II de notre article « Probability, Explanation, and Information », *Synthese*, vol. 48, 1981, p. 233-256.

²⁶ Une explication de ce type emploie un critère naturalisé lorsque la rationalité est définie en termes d'efficacité relative compte tenu des croyances et des désirs de l'agent. Un agent (plus ou moins) rationnel est donc quelqu'un qui est disposé à agir de manière (plus ou moins) efficace. Le fait de qualifier de « naturaliste » une telle explication comprend toutefois son lot de difficultés considérables, car les contraintes auxquelles sont soumises les attributions de croyances et de désirs au moyen d'un « principe de charité » peuvent compromettre l'idée suivant laquelle les explications qui recourent à la notion d'agent rationnel sont empiriques. Bien que je croie que cette difficulté puisse être surmontée, l'argument ne peut être présenté ici.

²⁷ Pour une vue d'ensemble de la littérature, cf. Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, Prentice-Hall, 1980. Il n'est pas surprenant que l'on y trouve une attention plus particulière accordée aux inconvénients qu'aux avantages.

²⁸ Pour rappeler un point de la section II : on peut procéder à des évaluations relativement à des critères évaluatifs particuliers sans pour autant établir ce qui les satisfait.

²⁹ Je suis conscient du fait que la caractérisation d'« un point de vue qui est à la fois impartial et prend en considération tous les individus qui pourraient être éventuellement concernés par cette évaluation morale » en tant que point de vue *social* peut induire en erreur – certains individus éventuellement concernés peuvent se trouver de l'autre côté d'une frontière intersociale. Cette complication sera mise de côté jusqu'à la section V.

³⁰ Cela signifie également que la relation entre les critères moraux et les critères de rationalité individuelle devient alors problématique puisque rien ne pourrait garantir que ce qui, du point de vue d'un quelconque individu, serait instrumentalement rationnel coïnciderait avec ce qui, d'un point de vue social, serait instrumentalement rationnel.

³¹ Une thèse plutôt forte à propos des comparaisons interpersonnelles est requise ici à des fins d'agrégation sociale. Je ne présuppose pas l'existence d'un seul bien, comme le bonheur [*happiness*], qui sous-tendrait de telles comparaisons. Ainsi, la théorie morale en question, bien que conséquentialiste, agrégative et maximisante, ne peut se réduire à l'utilitarisme classique. Je présuppose que lorsqu'il est question de faire un choix entre satisfaire un intérêt *X* de *A* par opposition au fait de satisfaire l'intérêt *Y* de *B*, les réponses à la question : « m'importerait-il davantage, toutes choses égales par ailleurs, d'avoir *X* satisfait si j'étais *A* plutôt que d'avoir *Y* satisfait si j'étais *B*? », seront relativement déterminées et stables parmi les individus dans des conditions où l'on dispose d'informations vives et complètes. Une forme similaire, mais un peu plus faible, de compatibilité parmi les différences est présupposée lorsque nous faisons des choix parmi des lignes d'action alternatives, qui nous conduiraient à avoir des désirs différents à l'avenir.

³² Voir par exemple, Barrington Moore, Jr., *The Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*, Boston, Beacon, 1966, et *Injustice: The Social Bases of Obedience and Revolt*, White Plains, N.Y., M. E. Sharpe, 1978; E. P. Thompson, *The Making of the English Working Class*, New York, Pantheon, 1963; William B. Taylor, *Drinking, Homicide, and Rebellion in Colonial Mexican Villages*, Stanford, Stanford University Press, 1979; Charles Tilly, *From Mobilization to Revolution*, Reading, Mass., Addison-Wesley, 1978; et Charles Tilly et al., *The Rebellious Century, 1830-1930*, Cambridge, Harvard University Press, 1975.

³³ Un thème commun que l'on retrouve dans les travaux cités à la note 32 consiste dans l'idée suivant laquelle une grande part d'agitation sociale est revendicative plutôt que révolutionnaire, étant donné que le mécontentement des groupes assaillis depuis longtemps mène très souvent à l'action et à la mobilisation seulement lorsque les droits coutumiers sont menacés ou niés. Ainsi, les idéologies manifestes de tels groupes sont fréquemment particularistes et conservatrices, tant même que leur agitation contribue à l'émergence de nouvelles formes sociales qui accordent un plus grand poids aux intérêts antérieurement mis à l'écart. De même, des individus qui ne parviennent pas à remarquer l'irrationalité de leur comportement coutumier jusqu'à ce qu'ils y soient conduits par la voie de difficultés non coutumières, ce qui éveille ensuite le sentiment que quelque chose ne va pas. Pour des raisons bien connues, une réponse individuelle initiale consiste généralement à renouer avec le *statu quo ante*, bien qu'il soit possible que des changements authentiques puissent avoir lieu à la suite de ces efforts de restauration.

³⁴ On doit souligner que ces mécanismes ne présupposent pas l'existence préalable d'institutions démocratiques. Ils ont pu extraire des concessions même au sein de sociétés qui sont demeurées profondément hiérarchiques. Voir, par exemple, Taylor, *Drinking, Homicide, and Rebellion*.

³⁵ En effet, le mécanisme décrit ici peut conduire simultanément à divers résultats : vers l'inclusion de certains intérêts antérieurement mis à l'écart, et vers l'exclusion de certains intérêts dont on tenait compte antérieurement. Pour en être certain, si des intérêts en viennent à être mis à l'écart même si leur base matérielle et sociale demeure plus ou moins intacte, un nouveau potentiel d'agitation sociale est créé. Certains groupes représentent alors un problème particulier en raison de leur incapacité inhérente à se mobiliser effectivement, tels que, par exemple, les enfants et les générations futures. Pour rendre compte des pressions qui ont été exercées au nom de ces groupes, il est nécessaire de rendre compte de la façon dont les individus en viennent à comprendre d'autres individus en fonction de leurs propres intérêts. (Comparons les façons par lesquelles nos diverses conceptions de nous-mêmes dans l'avenir, qui ne peuvent exercer aucune pression en elles-mêmes, en viennent à être prises en considération dans le moment présent en raison du fait que nous nous identifions avec elles.) À moins de tenir compte des processus d'incorporation et d'identification de ce genre, la moralité (et même la prudence) apparaîtra tout à fait mystérieuse, mais je ne me prononcerai pas davantage ici sur cette question. Pour quelques remarques préliminaires à ce sujet, cf. la section IX de mon article « Alienation, Consequentialism, and the Demands of Morality », *Philosophy and Public Affairs*, vol. 13, 1984, p. 134-171.

³⁶ Ici, tout comme ailleurs, nous entendons par « théorie morale contemporaine » les idées dominantes au sein des académies, et nous désignons par « discours moral contemporain » les pratiques répandues de l'argumentation publique en matière de moralité, généralement opérantes dans le cadre des sociétés où les sciences empiriques ont connu les plus hauts niveaux de développement. Rappelons de nouveau que le réalisme moral, tout comme le réalisme scientifique, ne s'engage pas à admettre l'idée d'un consensus global.

³⁷ Toutefois, de telles interdictions ont historiquement témoigné de limitations dans leur portée qui ne sont plus reconnues aujourd'hui comme valides. La tendance contre de telles limitations représente un exemple de la première tendance [*pattern*] évoqué ci-dessus, à savoir la tendance vers une généralité accrue.

³⁸ Qui plus est, les explications présentent un caractère très général, et opèrent à un niveau de description qui n'est pas à même de discriminer entre des hypothèses qui reposent sur la théorie particulière de ce qui est moralement bien [*rightness*], proposée ici, ainsi que d'autres théories qui se rapprochent plutôt de celle-ci. (Grosso modo, ceux qui caractérisent ce qui est bien [*rightness*] en termes de rationalité instrumentale, relative au bien non moral de celles et ceux concernés, mais qui diffèrent sur les détails de cette rationalité instrumentale – par exemple, est-elle simplement maximisante ou partiellement distributive? – ou du bien [*goodness*] non moral – par exemple, est-il réductible au plaisir? Pour une discussion de conceptions concurrentes et quelque peu divergentes, voir la section VI). Si on ne sert pas de la méthode employée ici pour faire des choix parmi des conceptions concurrentes et similaires, l'analyse empirique doit être alors beaucoup plus fine [*fine-grained*]. Des remarques similaires s'appliquent aux théories faibles et étendues de la rationalité individuelle, décrites ci-dessus.

³⁹ Il importe peu que nous affirmions qu'il n'ait *aucune* raison ou qu'il y a simplement une raison infiniment faible de poser cette action. Je suppose que nous pourrions même affirmer qu'une personne a une raison infiniment faible de faire quoi que ce soit – même de consacrer une quantité énorme d'efforts à purger des contradictions de ses croyances ou à purger des alligators d'un marécage lointain – qui pourrait de manière *concevable* s'avérer être à son avantage. Mais nous n'aurions alors aucune difficulté à garantir l'existence de raisons infiniment petites pour motiver la conduite morale. Cela permettrait, après tout, à une conception naturaliste du bien moral [*naturalized moral rightness*] de satisfaire à la thèse humienne de la practicalité, mais cette thèse ne serait pas pour autant dérobée de son intérêt.

⁴⁰ François (duc de) la Rochefoucauld, *Réflexions, ou sentences et maximes morales, suivies de réflexions diverses*, Éd. Jean Lafond, Paris, Gallimard, 1976, p. 79. La Rochefoucauld a apparemment emprunté cette phrase du clerc, Du Moulin. Je suis reconnaissant envers Barrington Moore Jr. de m'avoir rappelé cette citation. Cf. son ouvrage *Injustice*, p. 508.

⁴¹ Contrasteons ces idées avec le relativisme présenté par Harman au sujet des obligations [*ought*] dans *The Nature of Morality*. Harman adopte la première des deux stratégies présentées ci-dessus, de sorte qu'il conserve la relation entre les obligations morales d'un individu et la raison (instrumentale) qui motive sa conduite. Il défend son approche, en partie, en faisant valoir que si nous supposions qu'Hitler était engagé dans la poursuite rationnelle de ses fins, un jugement « interne » tel que « Hitler n'aurait pas dû (moralement) avoir tué 6 millions de juifs » serait « faible » ou « étrange » en comparaison avec un jugement « externe » tel que « Hitler était quelqu'un de mauvais [*evil*] » (voir, p. 107 sq.). Je serais porté à penser le contraire, à savoir qu'il est trop « faible » et « étrange » de formuler une théorie morale selon laquelle Hitler serait jugé comme étant quelqu'un de foncièrement mauvais (ce qu'Harman prétend accomplir avec son type de relativisme, sans présenter la moindre explication), mais dans laquelle 'Hitler (moralement) n'aurait pas dû agir comme il l'a fait' est faux.

⁴² Ce besoin ressenti se reflète également dans la codification des lois et dans le développement de théories légales. Bien que la loi puisse sembler parfois artificielle, les conditions sociales générales et les besoins qui ont motivé son développement sont pourtant bien réels. En effet, l'artifice élaboré de la loi et de son langage reflète en partie la nécessité pressante de surmonter le niveau de description du sens commun pré-théorique.

⁴³ Je suis redevable à plusieurs personnes, notamment Peter Achinstein, Robert Audi, Annette Baier, Michael Bratman, Stephen Darwall, Allan Gibbard, Thomas Nagel, Samuel Scheffler, Rebecca Scott, Nicholas Sturgeon, Nicholas White, et les éditeurs de *The Philosophical Review*, qui ont eu la gentillesse de me faire part de leurs commentaires sur des versions antérieures ou des présentations de cet article.