

RAYON PHILO

RUWEN OGIEN

Les causes
et
les raisons

philosophie analytique
et sciences humaines

ÉDITIONS JACQUELINE CHAMBON

Facebook : La culture ne s'hérite pas elle se conquiert

Les causes et les raisons

DANS LA MÊME COLLECTION

Jacques Bouveresse : *Langage, perception et réalité*
R. Casati et J. Dokic : *Essai sur la philosophie du son*
Donald Davidson : *Enquêtes sur la vérité et l'interprétation*
Jean-Yves Goffi : *Le philosophe et ses animaux*
J.-P. Leyvraz et K. Mulligan : *Wittgenstein analysé*
Claudine Tiercelin : *La pensée signe*

Couverture: Hugues Drouot

Copyright © 1995 by éditions Jacqueline Chambon, Nîmes

ISBN : 2-87711-136-9

ISSN : 1242-9066

Ruwen Ogien

Les causes et les raisons

Philosophie analytique et sciences humaines

ÉDITIONS JACQUELINE CHAMBON

AVANT-PROPOS

Après un demi-siècle ou plus de discussions sans concessions autour de l'« homme », du « sujet », des « structures », les sciences humaines se trouvent, en France, dans un état qui, de l'avis à peu près général, n'est guère reluisant.

On a l'impression qu'elles se sont trompées de sujet (si on peut dire) ou qu'elles se sont engagées dans un débat qui les dépassait complètement.

En mêlant leurs voix à celles des faux prophètes qui annonçaient, pêle-mêle, la mort de l'homme, de l'art, de l'éthique, de l'histoire, de la philosophie (qui ne s'en portent ni mieux ni plus mal, d'ailleurs), les spécialistes des sciences humaines ont fini par se retrouver sans voix, c'est-à-dire provincialisés, désarmés, quasiment incapables d'affronter un débat théorique ou épistémologique significatif.

Certes, il ne serait pas très difficile de comprendre la répulsion de ces spécialistes à l'égard des « débats théoriques », s'il était sûr que ces débats devaient obligatoirement ressembler à ce qu'ils étaient à la grande époque structuraliste, ou aux confrontations présentes entre rescapés de l'époque. Mais il est clair que ce n'est pas le cas.

Ainsi, dans le monde dit anglo-saxon, philosophes, logiciens, linguistes, psychologues n'ont jamais cessé de collaborer à une entreprise qui leur paraissait commune, féconde, cumulative, enrichie par les plus vifs débats et par la multiplication presque agaçante des théories. Et de leurs efforts croisés, de leur optimisme « naïf » est né, entre autres, ce qu'on importe aujourd'hui avec un enthousiasme suspect sous l'appellation plus ou moins contrôlée de sciences cognitives.

En un certain sens pourtant, les préoccupations de ces philosophes, logiciens, linguistes, psychologues, etc., n'étaient pas incommensurablement éloignées de celles qui, ici, empêchaient nos théoriciens de discuter sans finir par se battre. Ils se sont, eux aussi, intéressés à la « subjectivité », aux « structures ». Mais

ils l'ont fait à leur manière, sans pompe et sans grandiloquence, autour de thèmes modestes, accessibles à un ensemble très vaste de chercheurs.

Et ces discussions innombrables ont donné naissance à des courants théoriques importants qui traversent les disciplines les plus éloignées. C'est particulièrement évident dans le cas de la critique du behaviorisme, qui a mobilisé des philosophes, logiciens, linguistes, psychologues, sociologues trop nombreux pour être cités.

On peut avoir des raisons d'être accablé en considérant tout ce qui semble séparer, dans le style, le fond, la fécondité, ce débat de celui qui fut mené ici à partir de préoccupations voisines.

En même temps, on peut trouver quelques motifs d'espoir dans le fait que les sciences dites cognitives, issues de ce débat autour du behaviorisme en réalité, ont, semble-t-il, trouvé en France un public (ignorant probablement les origines modestes de ces sciences) et que, par leur biais, les sciences humaines regagnent un peu de leur vigueur argumentative, de leur unité, de leur faillibilité. Car l'un des défauts les plus rédhibitoires d'une théorie, c'est, bien entendu, l'immunité complète à l'égard de toute réfutation, et nos « grands théoriciens » donnent l'impression de ne s'être jamais posé la question de savoir s'ils pouvaient à la fois prétendre, comme ils le faisaient de façon souvent tapageuse, à la « scientificité », tout en ne donnant à leurs lecteurs aucun moyen de les réfuter.

À ceux qui trouveront ce diagnostic trop sombre, trop sommaire, déformé par les clichés dont on se sert habituellement pour opposer, par exemple, la philosophie « analytique » et la philosophie « continentale », on fera observer qu'il n'est, après tout, qu'une façon de prendre au sérieux la fameuse « exception culturelle française », que tant de bons esprits revendiquent (mais probablement pas en ce sens assez peu patriotique).

Cependant, si on souhaite vraiment tenir compte de ces différences « géophilosophiques » (un néologisme qui n'est ni très élégant ni très éclairant, mais qui a le grand avantage de ne pas être aussi compromettant que « culturel » ou « national »), on est bien obligé de compliquer un peu le tableau.

Il faudrait faire preuve d'une certaine étroitesse d'esprit pour ne pas prendre en considération les courants positivistes qui sont tou-

jours restés vivants en France et la tradition de « clarté cartésienne », qui, bien sûr, n'a jamais disparu, en dépit des déferlements d'hermétisme théorique (une tradition de clarté à laquelle on pourrait tout de même reprocher de se prendre un peu trop au sérieux, d'être une « clarté pompeuse », si on peut s'exprimer ainsi).

Inversement, il faudrait être aveugle ou malhonnête pour réduire le mouvement intellectuel dans le monde dit anglo-saxon à cette entreprise étroitement positiviste ou scientiste.

Ce mouvement intellectuel se distingue aussi par des critiques remarquables des tentatives d'aligner les sciences de l'homme sur les sciences de la nature.

Elles aboutissent à des conclusions qui ne sont guère éloignées de celles auxquelles les illustres critiques « continentaux » sont arrivés, dans le cours de la célèbre querelle sciences de l'esprit — sciences de la nature (ou comprendre-expliquer), conclusions qui sont approximativement les suivantes.

S'il est vrai que les sciences de l'homme ne peuvent pas s'aligner sur les sciences de la nature, elles sont confrontées à cette alternative :

— ou bien revendiquer la spécificité et l'autonomie des méthodes et des objectifs des sciences humaines, c'est-à-dire, faire admettre que le domaine de la scientificité déborde largement celui qui est circonscrit par les méthodes et les objectifs des sciences naturelles ;

— ou bien renoncer à revendiquer pour les sciences humaines quelque statut scientifique que ce soit, et les aligner sur l'art, le mythe, la vision du monde, ce qui signifie les affranchir des contraintes de la mesure, de la vérification, de la réfutation, mais pas nécessairement des tracasseries de l'évaluation. Car même si ce qu'on appelle sciences humaines s'apparente à l'art, au mythe, à la vision du monde, il n'en résulte pas que tout se vaut dans ces prétendues sciences. Certaines théories sont plus intéressantes, évocatrices que d'autres, de la même façon que certaines œuvres d'art sont plus séduisantes, fascinantes que d'autres et il n'est pas très facile de dire pourquoi.

L'alternative est bien connue et on trouvera certainement autant de partisans de l'un ou l'autre de ses termes en France et ailleurs. Mais la question n'est pas seulement de savoir si on admet l'alternative et si on est disposé à endosser l'une ou l'autre

des possibilités qu'elle offre. Elle est aussi de savoir pourquoi on l'admet et pourquoi on choisit l'un des termes de préférence à l'autre. Or, de ce point de vue, il faut, encore une fois, tenir compte de ces différences dites « géophilosophiques », qui devraient être anecdotiques mais qui sont loin de l'être, malheureusement.

Dans le monde dit anglo-saxon, la critique s'est concentrée autour de problèmes assez bien circonscrits : dichotomie des causes et des raisons, opposition entre les énoncés à la « première personne » et à la « troisième personne », débat autour de la possibilité de traduire, sans les altérer complètement, nos concepts mentaux (croyances, désirs, intention, conscience, etc.) en langage neuro-physiologique, discussion autour de la question de savoir si on peut proposer des explications satisfaisantes à un niveau dit supérieur (psychologique, sociologique, historique, etc.) ou si les seules explications authentiques sont de niveau dit inférieur (physique), et, en corollaire, question de la possibilité d'établir des lois strictes en sciences humaines ou d'attribuer aux événements mentaux (croyances, désirs, conscience) ou sociaux (degré d'intégration, taux de chômage, etc.) des pouvoirs causaux qui ne soient pas purement métaphoriques.

En France, la critique du scientisme ou du positivisme étroit ainsi que la défense de l'autonomie des sciences humaines fut, à un moment, une véritable industrie. Mais la cause était, semble-t-il, gagnée d'avance et les arguments ont plutôt servi de décoration. C'est vrai des arguments dits phénoménologiques qui ont perdu, à l'importation, toute leur précision pour devenir des lamentations nostalgiques sur l'homme, le « vécu », l'empathie. Et c'est vrai aussi des arguments dits épistémologiques, qui ont fleuri au bon temps (si on peut dire) du structuralisme, en changeant la rhétorique. Au vague désespérant des « phénoménologies », ils ont substitué la précision inutile, le formalisme d'apparat et l'hermétisme de principe, décourageant et intimidant.

Or c'est une chose de justifier l'autonomie des sciences humaines ou de nier leur existence au moyen d'intuitions vagues ou de propos délibérément hermétiques, qui ont en commun d'interdire la discussion, c'est est une autre, bien sûr, de le faire au moyen d'une discussion publique et argumentée, qui est loin d'être close, au demeurant.

Et ce n'est pas du tout la même chose d'exclure les sciences humaines du domaine de la science ou de leur accorder une autonomie d'objectifs et de méthodes après avoir pris la mesure du programme positiviste et de le faire par assurance dogmatique ou paresse, sans avoir consenti aux efforts nécessaires pour connaître les tentatives, parfois héroïques, d'intégrer les sciences humaines dans les programmes scientifiques forts. On pourrait même dire que la critique des anti-scientistes est d'autant plus fondée qu'elle prend appui sur les prétendus résultats des scientistes.

Ainsi, l'un des paradoxes de la recherche dans les sciences cognitives est que la réhabilitation de la vie intérieure (croyances, désirs, conscience, intentions, etc.) dont elle a fait le lit par la critique du behaviorisme, a servi de point d'appui pour nier la valeur de ces sciences. En effet, les cognitivistes ont réussi à montrer qu'on ne pouvait pas se passer, dans les sciences humaines, des notions de croyance, désir, représentation, intention, conscience, etc. Mais certains philosophes se sont appuyés sur cet « acquis » pour justifier leurs réticences à ranger les sciences humaines parmi les sciences ou même à leur accorder un statut scientifique autonome, parce que, d'après eux, aussitôt qu'on parle de croyance, de désir, de conscience, d'intention, il n'y a plus de place pour des lois, des hypothèses, des réfutations ; il n'y a même plus de place pour des « découvertes », de quelque ordre qu'elles soient. On peut ramener cet ensemble d'arguments à une sorte d'aporie.

À première vue, nous avons tout intérêt à nous débarrasser des notions intentionnelles (croyance, désir, raison, motifs, volonté, conscience, etc.) car elles sont, semble-t-il, les sources principales de nos difficultés dans l'explication de l'action humaine. Mais nous nous apercevons vite, en essayant de les éliminer, que nous n'avons aucune chance de réussir ! Les behavioristes et les physicalistes les plus optimistes finissent par le comprendre. En réalité, notre dilemme est le suivant : nos explications par les intentions, les croyances, les désirs, la conscience sont imparfaites, sources de confusions et de contradictions mais nous ne pouvons pas ou ne voulons pas y renoncer, pour toutes sortes de raisons, les raisons *morales* n'étant pas les moins importantes.

Tel est, du moins, le point de vue que j'aimerais pouvoir justifier dans ce livre. Mais, au-delà de cette aporie, c'est un certain

style de discussion que j'aimerais illustrer. Ce style, c'est celui de la philosophie dite analytique, dont le rôle dans l'unification des recherches positives en sciences humaines, aussi bien que dans la *critique* de ces sciences, est loin d'être négligeable.

Je n'ai pas du tout l'intention, bien sûr, de proposer des thèses précises sur un sujet aussi vaste et controversé que celui des relations sciences-philosophie. Je me contenterai de prendre acte du fait que la philosophie analytique n'a jamais interrompu son dialogue avec les sciences — contrairement à d'autres courants philosophiques — en dépit d'une tendance constante à l'autonomie, et en dépit, aussi, de tous ses avatars « post-analytiques ». Et j'essaierai simplement de montrer, au moyen de quelques exemples, ce que ce genre de réflexion philosophique peut apporter aux sciences humaines, sans préjuger du résultat auprès des spécialistes des disciplines traditionnelles, et sans exclure la possibilité que d'autres formes de réflexion — philosophiques ou non philosophiques — soient, elles aussi, en mesure d'apporter d'intéressantes contributions à la clarification de certains problèmes communs à l'ensemble de ces disciplines.

Ce livre rassemble une suite d'essais unis par la conviction que les sciences humaines en général ne perdront rien, si elles acceptent de renouer le dialogue avec la philosophie, lorsque ce dialogue s'est défait et lorsque cette philosophie est celle qu'on peut appeler analytique sans s'inquiéter de tout le flou qui entoure ce qualificatif. Je crois que ce serait un bon moyen de retrouver, lorsqu'il s'est perdu, le sentiment que la théorie n'est pas nécessairement hermétique ou grandiloquente et qu'elle n'est pas sans importance.

On pourrait exprimer cette conviction de façon plus laconique, en s'inspirant d'une formule bien connue : on a trouvé et on trouvera probablement encore des nuages de sciences humaines dans des gouttes de philosophie analytique.

INTRODUCTION

CE QU'ON NE PEUT PAS « NATURALISER »

Je n'ai jamais très bien compris pourquoi, dans les sciences dites humaines, des propositions dont la vérité ou la fausseté étaient manifestement établies indépendamment de toute enquête empirique étaient, néanmoins, présentées comme des sortes d'hypothèses plus ou moins confirmées par des données d'expérience.

Je ne fais pas allusion, bien sûr, à ces « explications causales » dont la circularité saute aux yeux (il n'est pas impossible de tomber sur des « découvertes » aussi creuses que : « la cause véritable du chômage c'est qu'il y a de plus en plus de gens qui n'ont pas d'emploi »). Je pense plutôt à des affirmations un peu moins vides (en apparence) telles que « l'injustice cause la souffrance morale » ou à des déclarations générales présentées comme des descriptions de faits naturels, telles que « la plupart des hommes sont égoïstes ou agissent de manière à maximiser leur bien-être ».

Il me semble que si ce genre de propositions sont réputées vraies, ce n'est pas ou ce ne peut être en raison du fait qu'elles sont plus ou moins bien confirmées par l'expérience. Et c'est en ce sens qu'on peut dire d'elles qu'elles sont *a priori*, puisque selon la définition minimale de *a priori* (du moins celles qui semblent faire l'objet d'un consensus minimal), les propositions dites *a priori* sont celles qui peuvent être tenues pour vraies, indépendamment de toute expérience (en un certain sens philosophique du mot expérience : fait d'observation directe ou rapporté, données de la mémoire ou même de l'introspection).

La question de savoir comment séparer, dans nos recherches, la partie dite non empirique ou *a priori* et la partie dite empirique ou *a posteriori*, ou plus exactement la question de savoir s'il est possible de les séparer dans nos recherches se pose dans toutes les disciplines et pas seulement dans les sciences humaines. Mais elle se pose avec une acuité particulière dans ces dernières.

Il peut ne pas être inutile d'insister, pour commencer, sur le fait que, lorsqu'on parle d'*a priori* en philosophie comme dans la plupart des réflexions théoriques dans les sciences spécialisées, ce n'est pas du tout en référence aux différentes formes d'ignorance volontaire, aux préjugés, aux superstitions, aux idées toutes faites, aux erreurs systématiques par absence de réflexion, d'attention, d'observation. L'*a priori* concerne des croyances vraies, des connaissances que nous aurions ou pourrions avoir par la simple force de nos raisonnements ou de nos intuitions non sensorielles, c'est-à-dire des connaissances qui ne dépendraient aucunement de l'exercice (correct) de nos facultés perceptives¹.

1. Étant donné le choc qu'a provoqué le célèbre article de Quine « Two Dogmas of Empiricism » (qui se proposait de ruiner la dichotomie analytique-synthétique et une certaine idée de l'*a priori*), il ne faut pas s'étonner de l'abondance de livres, articles, essais sur ces thèmes dans la philosophie anglo-saxonne (Quine W.V.O., « Two Dogmas of Empiricism », dans *From a Logical Point of View*, Cambridge, Mass., Harvard University Press, 1953, p. 20-46, trad. P. Jacob, « Les deux dogmes de l'empirisme », dans Jacob P., éd., *De Vienne à Cambridge*, Paris, Gallimard, 1980, p. 87-113.)

Les textes les plus intéressants et les plus controversés ont été rassemblés dans Moser P.K., *A Priori Knowledge*, Oxford, Oxford University Press, 1987. On y trouve entre autres : Lewis C.I., « A Pragmatic Conception of the *A Priori* (1923) », p. 15-25 ; Stroud B., « Wittgenstein and Logical Necessity » (1965), p. 68-84 ; Putnam H., « Analyticity and Apriority. Beyond Wittgenstein and Quine » (1979), p. 85-111 ; Chisholm R.M., « The Truths of Reason » (1977), p. 112-144 ; Kripke S.A., « *A Priori Knowledge, Necessity and Contingency* » (1980), p. 145-160. (Extrait de *Naming and Necessity*, 1972, Cambridge, Mass., University Press, trad. P. Jacob & F. Recanati, *La Logique des noms propres*, Paris, Minuit, 1982.) Parmi mes autres sources, je mentionnerai tout particulièrement : Bonjour L., *The Structure of Empirical Knowledge*, Cambridge, Mass., Harvard University Press, 1985 ; Baker, G.P., Hacker P.M.S., *Wittgenstein : Rules, Grammar and Necessity*, Oxford, Basil Blackwell, 1985 ; Bouveresse J., *La Force de la règle. Wittgenstein et l'invention de la nécessité*, Paris, Minuit, 1987 ; Proust J., *Questions de forme : logique et proposition analytique de Kant à Carnap*, Paris, Fayard, 1980.

Mais je suis déjà allé trop loin, ou trop vite, car je viens de proposer une définition plutôt subjective de l'*a priori*, c'est-à-dire que j'ai caractérisé l'*a priori* comme un certain mode de connaissance ou, dans un sens moins psychologique, comme un certain mode de justification de nos croyances. Or il existe d'autres définitions de l'*a priori*, qui ne sont ni psychologiques ni épistémologiques. C'est d'ailleurs pour ne pas réduire l'*a priori* à son sens épistémologique ou psychologique que des termes tels que analytique ou nécessaire ont été introduits. Mais pour éviter toute confusion, on choisit généralement de réserver le terme *a priori* aux formes de connaissance ou de justification des croyances, les termes analytique et nécessaire étant destinés, de leur côté, à qualifier des objets de connaissance (propositions, concepts, états de choses, etc.). Il en va de même, bien sûr, pour les antonymes : *a posteriori* est réservé aux formes de connaissance ou de justification des croyances ; synthétique ou contingent, aux objets de connaissance. En définitive, l'*a priori* conserve, assez souvent, son sens plutôt subjectif.

Les différences entre les aspects « subjectifs » ou « objectifs » de l'*a priori* sont loin d'être universellement admises. Il me semble, néanmoins, qu'il est utile de les garder à l'esprit, ne serait-ce que parce que ce sont souvent elles qu'on cherche indirectement à éliminer.

Il serait très difficile, par exemple, de comprendre pourquoi les positivistes logiques donnaient tant d'importance à la notion d'analyticité, si on ne tenait pas compte du fait qu'ils souhaitaient la réhabiliter dans l'intention, précisément, d'éviter toutes les absurdités dans lesquelles on risquait, d'après eux, de tomber lorsqu'on définissait l'*a priori* en termes subjectifs : appel à de mystérieuses facultés d'intuition non sensorielle, de perception directe d'idées claires et distinctes, etc.¹

Au lieu de s'attarder sur les états mentaux du sujet de la connaissance ou de faire des hypothèses étranges à propos de ses facultés mystérieuses, ils ont cru qu'il valait mieux s'intéresser à la structure objective des « vérités évidentes ». Ils ont pensé qu'il

1. Ashby R.W., « Linguistic Theory of the *A Priori* », dans *The Encyclopedia of Philosophy*, Edwards P., éd., New York, Free Press, MacMillan, 1967 ; Boghossian P.A., « Analyticity », dans Wright C., Hale B éd., *A Companion to the Philosophy of Language*, Oxford, Basil Blackwell, 1995.

n'était pas impossible de montrer qu'il ne s'agissait, au fond, que de tautologies, d'énoncés vrais par définition ou vrais en vertu de la signification des mots qui les composent, c'est-à-dire, dans leur vocabulaire, d'énoncés analytiques, par opposition aux énoncés synthétiques, dont la vérité ou la fausseté dépendrait de données d'expérience.

Mais si la thèse des positivistes logiques fut tellement discutée (et combattue) ce n'est pas parce qu'elle affirmait ce genre de principe assez vague et, au fond, assez inoffensif. Ce qui a paru extrêmement contestable, c'est leur tentative de réduire à de simples vérités logiques l'ensemble des énoncés dont la vérité ne paraît pas dépendre de données d'expérience.

Soit l'énoncé « Aucun célibataire n'est marié » ; c'est un exemple typique d'énoncé dont le caractère vrai, indubitable, ne semble aucunement dépendre de données d'expérience. Il serait absurde d'essayer de vérifier « Aucun célibataire n'est marié » en interrogeant tous les célibataires ou en observant leur comportement, alors que c'est exactement ce qu'il faudrait faire pour vérifier « Tous les célibataires sont ennuyeux », « Aucun célibataire n'est poli », etc. D'où vient, alors, le caractère apparemment indubitable de « Aucun célibataire n'est marié » ? La clé de l'énigme, c'est, pour le positiviste logique, le fait que les mots « non marié » et « célibataire » sont synonymes, ce qui nous autorise à les substituer l'un à l'autre. « Aucun célibataire n'est marié » ne veut rien dire d'autre que « Aucun non marié n'est marié ». Nous ne devrions, en principe, avoir aucune difficulté à saisir pourquoi ce dernier énoncé est toujours vrai ou vrai indépendamment de toute donnée d'expérience : c'est une sorte de tautologie. Mais Quine a établi (ou, du moins, a cru établir) dans un essai retentissant¹ que ce genre de transformation reposait entièrement sur une notion de synonymie confuse ou circulaire. Il en tirait la conclusion qu'on ne pouvait pas réduire ce genre d'énoncés dits analytiques à des vérités logiques (ou à de simples tautologies) et qu'il fallait chercher ailleurs la raison pour laquelle ils nous paraissaient évidents ou indubitables. Je ne rentrerai pas ici dans le détail de l'argument qui s'appuie sur une différence entre les substitutions légitimes dites extensionnelles, portant sur

1. Quine W.V.O., *op. cit.*, 1953.

deux façons différentes de désigner un même individu, et les substitutions illégitimes ou vouées à l'échec, consistant à essayer de substituer deux concepts, alors qu'ils sont faits précisément pour ne pas dire exactement la même chose. L'intérêt de l'argument, c'est qu'il devait contribuer à montrer que la division *a priori-a posteriori* ne pouvait certainement pas être sauvée en invoquant l'existence de vérités dites analytiques telles que : « Aucun célibataire n'est marié ».

Mais ce que Quine a surtout réussi à prouver, c'est que les frontières entre les énoncés dits analytiques et les énoncés dits synthétiques étaient loin d'être toujours clairement tracées.

Bien entendu, lorsqu'on nous assomme avec un énoncé tel que « Aucun célibataire n'est marié », il faut avoir l'ingéniosité de Quine pour montrer qu'il n'est pas analytique, en ce sens qu'il ne peut pas être vrai seulement en vertu de la « signification » des termes « célibataire » et « non marié ».

Mais considérons la liste d'énoncés suivante :

- Tous les cygnes sont des oiseaux.
- Tous les êtres humains sont mortels.
- Tous les cygnes sont blancs.
- Tous les êtres humains sont égoïstes.
- Tous les oncles sont des mâles.
- Tous les poissons vivent dans l'eau.

Il y a bien un cas où il semble qu'on ait affaire à ce genre d'énoncé vrai en vertu de la signification des termes (« Tous les oncles sont des mâles »). Mais dans tous les autres, la réponse est loin d'être évidente. On peut estimer par exemple que « être égoïste » est un trait essentiel de l'homme, ou un trait accidentel. Dans le premier cas l'énoncé est analytique, et dans le second il ne l'est pas.

Il n'est pas plus facile de s'appuyer sur l'idée de contradiction pour séparer nettement des classes d'énoncés analytiques et synthétiques, tout simplement parce que la notion de contradiction est aussi floue que celle d'énoncé analytique, dans la mesure où s'y trouvent confondues les impossibilités sémantiques, logiques et physiques. On peut se demander, par exemple, si « voir un son » est une impossibilité sémantique, logique ou physique. On peut même se poser la question de savoir si c'est vraiment une impossibilité.

Voici une liste de propositions apparemment absurdes mais dont il est difficile de détecter si elles le sont vraiment et pourquoi¹ :

- On peut voir ce qui n'existe pas.
- Il y a des douleurs inconscientes.
- Il y a des droites qui ne sont pas les plus courtes distances d'un point à un autre.
- Il y a des blocs de béton qui flottent sur l'eau.
- Il y a des sons qu'aucune créature ne peut entendre.

Au lieu d'opposer des énoncés vrais par définition ou vrais indépendamment de toute expérience et d'autres énoncés qui seraient vrais ou faux selon ce que dit l'expérience, Quine a proposé l'image d'une sorte de ligne continue, d'une différence de degré.

Il a estimé qu'il y avait, au centre de notre système de connaissance, un ensemble de croyances auxquelles nous ne voulons pas ou ne pouvons pas renoncer (les croyances logiques, par exemple) et, à la périphérie, ce qu'on appelle les énoncés contingents, que nous sommes toujours disposés à rejeter, avec, entre les deux extrêmes, toutes sortes d'états intermédiaires. Mais, en réalité, tout se tient et nos croyances logiques ou mathématiques, qu'on croit à l'abri de toute réfutation, sont, elles aussi, indirectement mises à l'épreuve par nos hypothèses physiques confirmées. Même nos croyances logiques sont, en principe, ouvertes à la révision.

On pourrait avoir l'impression que certains wittgensteiniens adoptent l'attitude de Quine puisqu'ils semblent estimer, eux aussi, que les divisions analytique-synthétique, *a priori-a posteriori* sont loin d'être clairement tracées. Ils ne semblent pas exclure la possibilité que la plupart des énoncés pourraient être traités ou bien comme analytiques ou bien comme synthétiques et toute connaissance ou bien comme valant *a priori* ou bien *a posteriori* ².

1. Pour d'autres exemples, Hoppers J., *An Introduction to Philosophical Analysis*, Londres, Routledge & Kegan Paul, 1990, p. 103-155.

2. Bouveresse J., *op. cit.* (1987) ; Schmitz F., *Wittgenstein, la philosophie et les mathématiques*, Paris, PUF, 1988 ; Blackburn S., *Essays in Quasi-Realism*, Oxford, Oxford University Press, 1993, p. 60-61.

Il n'y a rien, dans la forme linguistique d'un énoncé, par exemple, ou dans ce que nous pensons en général de nos connaissances qui nous obligerait à dire de tel énoncé qu'il est (intrinsèquement, en quelque sorte) analytique, de tel autre qu'il est (intrinsèquement) synthétique, de telle connaissance qu'elle est *a priori*, de telle autre qu'elle est *a posteriori*. En fait, si certains énoncés sont dits analytiques et certaines connaissances *a priori*, c'est seulement parce que nous avons décidé de les traiter comme tels.

Ainsi l'énoncé « Tous les hommes sont égoïstes » peut-être traité comme une hypothèse réfutable. Mais il n'est pas du tout impossible de lui faire jouer le rôle d'un énoncé dégagé, en grande partie, de ses responsabilités à l'égard de l'expérience. C'est bien ce qui se passe lorsqu'il prend la place d'un principe premier dans la théorie de la décision ou dans certaines théories économiques classiques. L'énoncé est formellement (ou linguistiquement) le même, mais notre attitude à son égard a radicalement changé. L'énoncé est pour ainsi dire « classé dans les archives », soustrait aux aléas de la confrontation directe avec l'expérience, « ennobli ».

Il est possible que toutes sortes de considérations, instrumentales ou autres, aient joué un rôle dans la décision d'« ennoblir » l'énoncé, de le faire passer de son modeste statut descriptif à ce rôle directeur ou normatif. Mais, comme pour tout « ennoblissement », les origines modestes sont, en quelque sorte, oubliées (ou tout est fait pour qu'elles le soient). De toute façon, la connaissance des origines des énoncés normatifs ne peut, en aucune façon, nous aider à comprendre le rôle qu'ils jouent en tant qu'énoncés normatifs.

Aucun de ces arguments ne semblent contredire l'idée qu'il n'est pas déraisonnable d'adopter une attitude plutôt pragmatique à l'égard des divisions *a priori-a posteriori*, analytique-synthétique, c'est-à-dire de considérer qu'elles sont, tout simplement, le produit de décisions plutôt qu'une expression de contraintes « objectives » (sémantiques ou autres).

De ce point de vue, il y a bien quelque chose en commun entre les positions de Quine et celles des wittgensteiniens. Mais il y a néanmoins d'importantes différences entre elles, dans la mesure où les seules considérations qui semblent présider à nos choix sont, d'après Quine, instrumentales et dirigées vers l'accroisse-

ment de nos connaissances scientifiques, alors que pour un wittgensteinien, des considérations non instrumentales peuvent aussi entrer en ligne de compte.

Cependant, dans leurs versions instrumentales aussi bien que dans leurs versions non étroitement instrumentales, ces thèses ont été critiquées pour leur saveur exagérément conventionnaliste. Elles semblent dire, en effet, que c'est nous qui décidons, de façon assez arbitraire en définitive, ce qui compte comme possible ou impossible, contingent ou nécessaire. Or, dire qu'on décide librement, volontairement par proclamation, convention, accord, etc., de déterminer ce que sont les nécessités, les possibles et les impossibles, logiques et physiques, c'est vraiment aller de façon gratuite à l'encontre de nos intuitions les mieux enracinées¹. Il est difficile d'admettre que les règles logiques pourraient être modifiées selon notre bon plaisir (même ceux qui

1. Dans ses essais sur le quasi-réalisme, Simon Blackburn a résumé, indirectement, les objections de ceux qui ne supportent pas la « désinvolture » conventionnaliste : « Nous admettons des possibilités, nous excluons des impossibilités, nous insistons sur des nécessités. Dans aucun de ces cas nous ne décrivons quelque chose. Comme le dirait Wittgenstein, attribuer la nécessité à une proposition, ce n'est pas dire quelque chose de vrai ou de faux de cette proposition, ou du moins ce n'est pas ainsi qu'il faut comprendre cette attribution. Attribuer de la nécessité, c'est plutôt comme adopter une norme, ou une politique ou une règle disant qu'on classe une thèse dans les archives, au-dessus du tohu-bohu des faits empiriques. Cette décision gouverne la façon dont nous traiterons les faits récalcitrants. Tout cela s'accorde très bien avec le parallèle [entre morale et modalité]. Un genre de règle rend certaines pensées intellectuellement obligatoires, un autre, certaines actions. Néanmoins, un problème considérable se pose : identifier l'espace de jeu dans la production de ces règles. Modaliser ne semble pas plus être une affaire de choix que moraliser : il semble que nous considérons que "1 + 1 = 2" est nécessaire, tout simplement parce que nous *devons* le faire, et non parce que nous avons choisi de le faire. Le statut de "1 + 1 = 2" est plus naturellement perçu comme un produit de notre incapacité de concevoir autre chose, ou de *faire* quoi que ce soit avec des jugements contre-arithmétiques. Même si c'est nous qui conférons la nécessité, en quelque sens que ce soit, il reste qu'il serait assez peu naturel de la concevoir comme reflétant quoi que soit que nous pourrions choisir. » (Blackburn S., *op. cit.*, 1993, p. 60-61.)

pourraient trouver beaucoup d'avantages à l'admettre auront, semble-t-il, les plus grandes difficultés à se convaincre qu'une telle possibilité existe réellement). Il est assez évident aussi qu'on ne peut pas changer les règles logiques comme on peut changer les conventions qui règlent la circulation.

Cependant, dire que c'est nous qui faisons le partage entre ce que nous tenons pour synthétique ou *a posteriori* et ce que nous tenons pour analytique ou *a priori* ne signifie pas que nous le faisons de façon capricieuse ou parfaitement arbitraire.

On a vu que Quine semblait penser que des considérations instrumentales limitaient ces choix. À la suite des wittgensteinien, on peut envisager d'autres limites, non étroitement instrumentales. Pour faire droit à toutes ces possibilités, nous devons laisser ouverte la question de savoir pourquoi, par exemple, nous décidons que des affirmations telles que « Tous les hommes agissent pour maximiser leur bien-être » ou « Tout le monde vise le bonheur » doivent être considérées comme des vérités inattaquables. Nous les avons promues à ce rang pour des raisons qui n'ont peut-être rien à voir avec les progrès instrumentaux de notre connaissance ou avec la possibilité d'une maîtrise accrue de notre environnement. C'est à partir de questions de ce genre que je m'efforce de reconsidérer l'analyse de Davidson, et son traitement de l'esprit et de l'action.

2

La manière la plus concise ou la plus commode de présenter mon argument général serait de dire que j'adopte l'attitude interprétationniste de Davidson, mais en insistant sur le contenu moral de son principe de charité¹.

En réalité, je ne propose rien de plus qu'un amendement (suffisant, je l'espère, pour mériter un peu d'attention !) aux théories interprétationnistes à la Davidson².

1. Davidson D., *Actions et événements* (1982), trad. P. Engel, Paris, PUF, 1993.

2. Je ne suis pas le seul, bien sûr, à proposer un amendement de ce genre. Il a été suggéré notamment par Susan Hurley, dans *Natural Reasons*, Oxford, Oxford University Press, 1989, p. 84-101 ; et aussi, bien sûr, par Ronald Dworkin, *A Matter of Principle*, Cambridge, Mass., Harvard

Il me paraît difficile de nier que, dans nos tentatives de comprendre autrui et nous-mêmes, nous visons la meilleure interprétation rationnelle possible. Mais j'estime que celle-ci est, ou doit être, une interprétation morale, c'est-à-dire une interprétation rationnelle non étroitement instrumentale.

Proposer la meilleure interprétation rationnelle possible, c'est appliquer ce que Davidson appelle un principe de charité. Ce principe peut être décrit comme un ensemble de recommandations : considérez que votre interlocuteur comprend ce qu'il dit, respecte le principe de contradiction, essaie de mettre ses actions en harmonie avec ses désirs et ses croyances, que ses croyances sont en grande partie vraies, qu'elles correspondent aux vôtres ; n'attribuez l'infraction à certaines règles rationnelles qu'en dernier recours et l'abandon de la rationalité minimale que dans les cas désespérés ; et même dans ces cas, demandez-vous, avant de fermer le dossier, si ce n'est pas *vous* qui êtes un détestable interprète.

Tout ceci serait parfaitement convaincant, si on pouvait exclure la possibilité de donner à l'idée de « meilleure interprétation rationnelle possible » une signification purement instrumentale, c'est-à-dire exclure la possibilité que cette meilleure interprétation rationnelle consiste seulement à proposer une image cohérente relativement à des buts, motifs, intentions, quels qu'ils soient.

C'est que, dans la perception que nous avons d'autrui ou de nous-même en tant qu'êtres rationnels, ce qui compte, ce n'est pas seulement que nous puissions avoir des raisons de penser que les moyens les mieux appropriés ont été mis au service de fins quelconques. Le genre de vie qu'on veut mener, le genre de personne qu'on veut être, la nature et la qualité objective des buts qu'on se propose font partie de ce tableau rationnel.

Il me paraît assez improbable qu'une personne ordinaire jugera « rationnel » celui qui, comme le dit Hume, préfère la

...University Press, 1985, p. 167-177. Mais Robert Nozick a donné d'assez bons arguments contre leurs formulations dans *The Nature of Rationality*, Princeton, Princeton University Press, 1993, p. 133-181 ; et j'essaie de justifier cette distinction sur d'autres bases, en exploitant par exemple la théorie pragmatique de *l'a priori*.

destruction du monde entier à une égratignure de son doigt¹. En définitive, un interprétationniste devrait dire de cet extravagant personnage de Hume qu'on ne doit pas le croire, ou qu'on le croira peut-être si on a épuisé toutes les autres hypothèses qui pourraient rendre ses déclarations un peu moins absurdes.

Bref, il me semble que dans l'interprétation, ce que nous essayons de construire, c'est la meilleure image morale possible d'autrui et de nous-mêmes, parce que c'est, en un sens, la meilleure image rationnelle qu'on puisse proposer. Ce qui n'exclut pas, bien sûr, que nous ne puissions pas faire autrement, souvent, que diagnostiquer le mal ou l'irrationalité.

Je ne pense pas que l'argument de Davidson, considéré dans toutes ses implications, pourrait admettre une interprétation purement instrumentale. Cependant Davidson ne donne pas toujours l'impression d'exclure une interprétation de ce genre, en particulier lorsqu'il recommande d'examiner le problème de la faiblesse de la volonté indépendamment de toute considération « morale ». C'est pourquoi je préfère préciser et parler de *meilleure interprétation morale possible*.

Bien entendu, on pourrait m'objecter que l'expression « meilleure interprétation morale possible » est extrêmement vague : un kantien ne verra pas les choses de la même façon qu'un aristotélicien ou un utilitariste, etc.

Mais dire qu'on cherche la meilleure interprétation rationnelle ou morale possible ne signifie pas qu'on ait des idées définitives sur le contenu de ces interprétations².

Il est possible que cette réponse à l'objection donne l'impression de n'être qu'une façon d'éviter l'objection sans y répondre. Pour-

1. Hume D., *Traité de la nature humaine*, 1739-1740, trad. Leroy A., Paris, Aubier, 1946, p.525.

2. Ce que dit Pascal Engel sur cette question est particulièrement éclairant. Cf. « Trois formes de normativité », dans *Lire Davidson. Interprétation et holisme*, Combas, L'éclat, 1994 a, p. 205-225. Sur la question de l'interprétation en général (et sur beaucoup d'autres discutées ici) cf. aussi Dan Sperber, *Le savoir des anthropologues*, Paris, Hermann, 1982 ; cf. également, Patrick Pharo, *Le sens de l'action et la compréhension d'autrui*, Paris, L'Harmattan, 1993.

tant, il est difficile d'en proposer une autre, à partir du moment où on choisit d'adopter ou d'exploiter, comme je le fais, une notion d'interprétation à la Davidson.

En effet, il y a deux façons (au moins) de concevoir l'interprétation : comme une procédure régulatrice et comme une procédure constitutive.

Lorsque l'interprétation est conçue comme une procédure régulatrice, elle ressemble à un travail inductif. Interpréter, ce serait organiser des faits et des données d'une certaine façon ; ce serait, en quelque sorte, proposer une hypothèse particulière.

Mais l'interprétation constitutive n'est pas du tout une procédure inductive. Elle ne dépend pas de faits ou de données : elle les fait apparaître. Ce n'est pas une hypothèse particulière mais une sorte de présupposé extrêmement général.

L'interprétation à la Davidson me semble plutôt constitutive que régulatrice. Et, de ce fait, elle reste très générale : c'est un ensemble de principes qui orientent toute enquête plutôt qu'une hypothèse portant sur des faits déterminés.

On me dira peut-être à présent que, même si cette signification constitutive du terme interprétation est acceptable, il n'en résulte pas qu'elle soit applicable aux états mentaux et aux comportements humains, qui semblent être des faits ou des données qu'on doit appréhender en formulant des hypothèses. Bref, il reste encore à montrer qu'une position interprétative constitutive est acceptable en philosophie de l'esprit et de l'action. De très nombreux et très bons arguments ont été proposés en faveur de l'interprétation constitutive des états mentaux. Ils ne coïncident pas exactement avec ceux qui portent sur l'action.

Les bons arguments en faveur de l'interprétation constitutive en philosophie de l'esprit sont des arguments sceptiques ou des thèses adoptées par défaut. Ces arguments disent qu'il n'y a aucune raison décisive d'être réaliste pour les états mentaux. Les bons arguments en faveur de l'interprétationnisme en philosophie de l'action sont un peu moins négatifs. J'aurai l'occasion d'évoquer les arguments interprétationnistes en philosophie de l'esprit. Dans cette introduction, je me contenterai de présenter, de façon très générale, ceux que je défendrai tout au long de ce livre, qui porte plus précisément sur la philosophie de l'action et ce qu'elle nous apprend, ou peut nous apprendre, des limites de ce qu'on appelle les sciences humaines.

Il existe trois difficultés qui, à mon sens, légitiment l'interprétationnisme en philosophie de l'action :

- 1) le problème des chaînes causales déviantes ;
- 2) les mystères de l'agentivité ;
- 3) les limites des théories qualitatives de l'action.

Commençons par les chaînes causales déviantes.

Si tout ce qu'on veut dire en parlant de relation causale entre des désirs des croyances et des mouvements corporels, c'est qu'on peut observer certaines régularités dans l'apparition de certains événements physiques et mentaux, le problème des chaînes causales déviantes se posera inévitablement.

Il peut y avoir coïncidence entre des désirs et des croyances préalables et l'état de choses qui s'est réalisé, sans que celui-ci se soit réalisé à cause de ces désirs et de ces croyances. C'est exactement ce qui se passe dans certains films policiers : on y voit des criminels déçus parce que leur victime est déjà morte lorsqu'ils tombent dessus. Mais les cas qui sont évoqués dans la discussion des chaînes causales déviantes sont un peu plus complexes¹.

On envisage, la plupart du temps, le cas d'une personne qui veut tuer quelqu'un et qui fait quelque chose dans cette intention, comme lui tirer dessus, ou l'écraser avec sa voiture. À la fin de l'histoire, la victime est bien morte, mais entre l'initiative et la réalisation de la fin souhaitée par l'assassin, il s'est produit quelque chose qui a été causé, certes, par cette initiative, mais qui n'était pas sous le contrôle de l'assassin. Par exemple, il tire, il rate sa victime, mais la balle frappe un curieux penché à sa fenêtre qui tombe sur la tête de la victime et la tue. Peut-on dire que la mort de la victime doit être entièrement imputée au tireur ? Est-ce que le tireur a eu de la chance ou de la malchance ? Il existe probablement de bonnes réponses juridiques à ces questions mais je ne suis pas sûr qu'elles soient philosophiquement satisfaisantes. Quoi qu'il en soit, ce genre d'exemple donne des raisons d'estimer que la correspondance entre l'antécédent et le conséquent ne suffit pas et qu'il faut avoir certaines garanties

1. Searle J., *L'Intentionnalité : essai de philosophie des états mentaux* (1983), trad. C. Pichevin, Paris, Minuit, 1985, p. 136-139.

d'agentivité pour le processus tout entier qui ne serait une action que si, et seulement si, il pouvait être imputé à l'agent du début à la fin, sans résidu¹. Mais c'est une condition qu'il est impossible de satisfaire, entre autres, parce que le problème des chaînes causales déviantes se pose non seulement pour ce qu'on pourrait appeler l'intention générale mais aussi pour chacune des intentions particulières ou des sous-intentions.

Entendue au sens d'une intention préalable, l'intentionnalité ne garantit pas que l'action puisse être mise au compte de l'agent, parce qu'entre cette intention préalable et l'état de choses final, il peut n'y avoir qu'une relation de coïncidence. Entendue au sens d'un contrôle constant du processus allant du plan initial à la réalisation de l'état de choses final, l'intentionnalité est une chimère. On ne peut pas exclure la possibilité que la maîtrise de certaines parties du processus, au moins, ait échappé à l'agent. On peut, certes, laisser à l'agent lui-même le soin de décider si cette action est la sienne en dépit de sa maîtrise partielle du processus. Mais, dans ce cas, l'assimilation du concept d'intentionnalité à celui de causalité par les désirs et les croyances ne tient plus. En fait, on passe d'une version *réaliste* à une version *interprétationniste* de l'intentionnalité, car si l'agent peut décider qu'une action est la sienne en fonction de considérations qui ne sont pas liées à ce qui s'est produit effectivement dans la chaîne des événements, rien n'interdit à un observateur-interprète de le faire à sa place, conformément à certaines considérations morales, sociales ou juridiques.

La conception téléologique de l'action n'échappe pas à ce dilemme². Ou bien on estime que l'agent contrôle effectivement toutes les séquences qui doivent contribuer à la réalisation de la fin vers laquelle il tend et corrige effectivement toutes les déviations de trajectoires, ce qui, dans le cas des actions humaines est parfaitement irréaliste ou indémontrable ; ou bien on pense que l'agent donne un sens téléologique global à son action, en décidant qu'elle est la sienne du moment que son but a été réalisé, même s'il n'a pas contrôlé ou corrigé effectivement toutes les

1. *Id.*, p. 116-120.

2. Taylor C., *The Explanation of Behavior*, London, Routledge & Kegan Paul, 1964.

séquences, ce qui revient à endosser une théorie interprétationniste. De toute façon, la version réaliste de l'action téléologique n'est pas causale à proprement parler car elle suppose l'existence et l'intervention d'un agent visant des fins, personnage dont les causalistes stricts préfèrent se passer mais qu'il est, tout compte fait, impossible d'éliminer. On en vient donc, assez natufellement, à se poser la question de l'agentivité.

Les causalistes forts ou stricts réservent l'expression « intentionnalité » à la description de relations entre événements mentaux et événements physiques, que les premiers soient ou non accompagnés de conscience, qu'ils soient attribués à un agent ou pas. Il y a action lorsque la flèche de la causalité va des événements mentaux aux événements physiques.

Mais si l'action se résume à cet ensemble de relations causales entre événements qui se produisent en moi ou qui m'arrivent, comme si je n'en étais que le réceptacle ou le spectateur impuissant, pourquoi l'appeler action ? N'est-ce pas, en réalité, une succession d'événements subis ?

Par ailleurs, rien ne nous interdit d'envisager la possibilité que les désirs et les croyances soient, en quelque sorte, sélectionnés ou appréciés non seulement de l'extérieur mais aussi de l'intérieur. C'est bien cette possibilité qu'on évoque lorsqu'on affirme que les croyances et les désirs peuvent être acceptés, rejetés, évalués, maîtrisés¹. Mais *qui* procède à l'évaluation des désirs et des croyances ? *Qui* est tantôt fier tantôt honteux de les avoir ?

Ceux que ces problèmes embarrassent se sont demandé s'il ne valait pas mieux substituer à l'idée d'une causalité des événements celle d'une causalité de l'agent². Mais la conception qu'ils se font de l'agent est parfaitement mystérieuse. L'agent serait, pour ainsi dire, situé en dehors de ses désirs et de ses croyances.

1. Taylor C., « What is Human Agency », dans *Philosophical Papers*, I, *Human Agency and Language*, Cambridge, Cambridge University Press, 1985, p. 17-49 ; Frankfurt H., « La liberté de la volonté et la notion de personne », dans Neuberger M., éd., *Théorie de l'action. Textes majeurs de la philosophie analytique de l'action*, Bruxelles, Mardaga, 1991, p. 253-269.

2. Chisholm R.M., *Person and Object*, Londres, Allen & Unwin, 1976, p. 69-72.

Mais à quoi pourrait bien ressembler un agent indépendant de ses propres désirs et croyances et capable, de ce fait, de les sélectionner ou de les causer ?

Heureusement, ce problème ne se pose de façon aiguë que si l'agent est censé intervenir en tant que cause réelle de l'action, en tant qu'objet du monde ou appartenant au monde au même titre que d'autres objets. Il disparaît lorsque l'agent est introduit dans la description en tant que fiction morale ou juridique, ou en tant que limite du monde, un peu comme s'il était un point sans étendue¹. Passons aux limites des théories qualitatives de l'action.

Certains philosophes pensent que, si les causalistes événementiels sont incapables de résoudre la question des chaînes causales déviantes ou de dissiper les mystères de l'agentivité, c'est parce qu'ils ne tiennent aucun compte de la dimension qualitative de l'agir, c'est-à-dire de ce que cela « fait » d'agir, de ce qu'on « éprouve » en agissant, etc. Les philosophes estiment que les énigmes de l'action pourraient bien se dissoudre, si on revenait aux notions assez ordinaires de conscience d'agir, sentiment de l'effort, expérience de l'action, etc. Une partie d'entre eux ont même cru trouver dans cette dimension qualitative des arguments en faveur de la possibilité de « naturaliser » l'action.

D'après Searle, par exemple, il semblerait qu'on puisse établir le fait qu'il existe une différence dans ce que l'agent « ressent » lorsqu'il lève la main de son plein gré et lorsqu'il le fait sous l'influence d'une manipulation chimique ou électrique, de sorte que ce que l'agent « ressent » pourrait servir de critère de distinction entre ce qu'il fait et ce qui lui arrive². Mais, à supposer que ce genre d'hypothèse soit vérifiable et plus ou moins confirmée, elle ne serait pas très instructive. Elle ne nous donnerait certainement pas les moyens de distinguer l'action de l'événement subi dans tous les cas qui nous préoccupent habituellement. Il serait assez absurde d'essayer de vérifier une hypothèse

1. Hart H.L.A., « The Ascription of Responsibility and Rights », *Proceedings of The Aristotelian Society*, 49, 1948, p. 171-194 ; Anscombe G.E.M., « The First Person », dans Guttenplan S., éd., *Mind and Language*, Oxford, Clarendon Press, 1975, p. 45-65.

2. Searle J., *op. cit.*, 1985, p. 114-115.

disant qu'il y a une différence « qualitative » évidente entre lever la main de son plein gré pour enlever son chapeau quand il fait trop chaud et lever la main pour enlever son chapeau parce qu'on nous a donné l'ordre de le faire (à l'armée, en prison, dans un lieu de culte, etc.).

Cependant, ce qu'on appelle « expérience de l'action » ne se résume pas à ce qu'on peut « ressentir ». Il y a des définitions non qualitatives de cette expérience, celles, en particulier, qui l'assimilent à un savoir pratique, un savoir-faire. C'est en ce sens qu'on peut interpréter, semble-t-il, le concept de connaissance sans observation, proposé par Anscombe¹ pour caractériser l'expérience de l'action. Au demeurant, les tests utilisés pour identifier la perspective de la première personne ne sont pas nécessairement qualitatifs. On peut, en effet, caractériser les phrases d'action à la première personne à partir de :

1) l'immunité à l'égard des erreurs d'identification (il est exclu que l'agent ait à se demander dans des conditions normales : « Quelqu'un croise les jambes. Est-ce moi ? »)²

2) la position d'autorité relativement aux affirmations portant sur les intentions, bruts, visées de l'action (on ne peut pas corriger les déclarations de l'agent à propos de ce qu'il essaie de faire, ce qui prouverait, d'après Anscombe, que ce genre de déclaration n'est pas descriptive puisque toute description peut être corrigée)³.

Dans l'ensemble des théories de l'action à la première personne, il convient de distinguer les théories purement qualitatives qui mettent l'accent sur la conscience, les sensations de mouvement, le sentiment de l'effort, etc. et les théories non qualitatives qui insistent plutôt sur la position objective exceptionnelle de l'agent relativement à ses actions. Or, c'est plutôt du côté des théories qualitatives qu'on cherche des arguments en faveur de la naturalisation de l'action et la faiblesse de ces théories qualitatives atteint, bien entendu, le projet de « naturalisation ».

1. Anscombe G.E.M., *Intention*, Oxford, Basil Blackwell, 1957, p. 14.

2. Evans G., *The Varieties of Reference*, Oxford, Oxford University Press, 1982, p. 215-220.

3. Anscombe G.E.M., *op. cit.*, 1957, p. 14.

Je ne crois pas aller trop loin en suggérant que les théories de l'action à la première personne non qualitatives sont, en réalité, des théories interprétatives.

Pour finir, je vais essayer de me montrer un peu moins prudent. En développant mes arguments sur les mystères de l'agentivité j'ai peut-être donné l'impression d'admettre que le problème de la liberté ou de la réflexivité ne se posait pas au plan des intentions, mais au plan du *choix* des intentions, dont la sélection semble dépendre de l'existence d'un agent. Mais, en réalité, même l'intention première, si on peut dire (ou l'intention de premier ordre), doit être conçue comme une intention libre. Tel sera, du moins, mon dernier argument en faveur des théories interprétationnistes de l'action.

Pour montrer à quel point nos idées de liberté et d'intention sont liées, je commencerai par proposer un argument qui a été souvent utilisé contre les tentatives de donner une explication biologique ou des fondements « naturels » à l'éthique¹. La crédibilité de ces explications repose, en grande partie, sur le fait que différentes espèces animales semblent se comporter naturellement d'une façon dite altruiste, c'est-à-dire qu'elles montrent des comportements de sacrifice individuel au bénéfice du groupe ou de l'espèce, ainsi que des comportements de coopération dont les bénéfices individuels sont loin d'être évidents. Mais on a objecté qu'il était absurde d'appeler altruistes ces comportements de sacrifice ou de coopération, dans la mesure où notre notion courante d'altruisme n'est pas purement ou étroitement comportementale. Qu'est-ce que cela veut dire ?

Eh bien, si nous estimons qu'un comportement de sacrifice ou de coopération apparemment dépourvu de bénéfice personnel n'est, en réalité, qu'une invitation à la réciprocité stricte ou un investissement social ou génétique à long terme, nous n'aurons plus de raison de l'appeler altruiste. C'est que, dans nos façons d'identifier l'altruisme, les intentions comptent. Et comme nous n'avons aucune idée de ce que peuvent être les intentions des fourmis, par exemple, nous dirons que les naturalistes utilisent le terme « altruisme » ou bien de façon naïve ou bien de façon plutôt métaphorique.

1. Changeux J.P., éd., *Fondements naturels de l'éthique*, Paris, Odile Jacob, 1991.

Mais que se passerait-il si nous apprenions que nos fourmis possèdent en effet quelque chose qui ressemble vaguement à des intentions, mais qu'elles ne peuvent, de toute façon, avoir que ces intentions-là, et qu'elles ne peuvent rien faire d'autre que ce qu'elles font ? Il me semble que cela suffirait (ou devrait suffire) à nous interdire de parler d'intention à leur propos, car notre notion d'intention est liée, en un sens apparemment très important, à celle de liberté d'agir, ainsi qu'à la notion de responsabilité, comme le montrent suffisamment nos pratiques judiciaires.

Je ne vais pas approfondir cette question ici. Je me contenterai d'envisager ce qui devrait suivre dans nos façons de traiter la question de l'intentionnalité si on admettait qu'elle est étroitement liée aux idées de liberté et de responsabilité. Il me semble qu'il faudrait reconnaître, alors, que la question de l'existence des intentions n'est peut-être pas plus une question de fait que la question de l'existence de la liberté.

De même qu'on peut séparer la question du déterminisme naturel de celle de la liberté humaine, en ce sens, que ni les arguments déterministes ni les arguments indéterministes ne sont décisifs pour trancher la question de la liberté, de même, me semble-t-il, on peut séparer la question de l'existence des intentions de tout ce que la neuropsychologie et la neurophysiologie peuvent ou pourraient nous apprendre, en ce sens que les arguments naturalistes *contre* la réalité des intentions ne sont pas plus décisifs que les arguments déterministes *pour* la réalité des intentions.

Ce n'est pas sur le terrain des faits physiques-cérébraux ou autres que se tranche la question de l'existence des intentions, pas plus que ce n'est sur le terrain des faits naturels que se tranche la question de la liberté, pour la bonne raison que ces questions sont, en réalité, du même ordre. Elles portent sur des présupposés, des normes qui orientent toute enquête, et qui, en un certain sens, ne peuvent être réfutés ni confirmés par quelque fait que ce soit.

En résumé, j'espère pouvoir montrer que trois difficultés dans les théories de l'action rendent l'approche interprétationniste non naturaliste à peu près inévitable : les chaînes causales déviantes ; les mystères de l'agentivité ; les limites des théories

qualitatives de l'action. Je me suis permis d'ajouter une difficulté supplémentaire en appendice (au cas où les trois autres ne suffiraient pas !) : le parallélisme entre la question de l'intention et celle de la liberté.

4

Les deux premières parties de ce livre s'adressent à l'une ou l'autre de ces difficultés. Et la morale qu'on devrait pouvoir tirer de ces discussions dirait qu'il est bien difficile (pour ne pas dire impossible) de naturaliser l'action. Dans la troisième partie, j'essaierai de montrer aussi pourquoi il est si difficile (pour ne pas dire impossible) de naturaliser l'éthique.

Étant donné que je donne un sens assez étroit (ou non métaphysique) au terme naturalisation (naturaliser, ce n'est rien d'autre que « réduire » ou « éliminer » toutes les explications non physicalistes, ou plus exactement, non formulées dans les termes des théories physiques existantes), cet ensemble d'objections à la naturalisation n'est pas du tout une invitation (qui serait absurde) à renoncer aux sciences humaines.

C'est même le contraire qu'il faudrait dire : montrer les limites de la naturalisation revient, en réalité, à préserver la légitimité des explications non purement physicalistes et, en définitive, à sauver, pourrait-on dire, un certain nombre de théories sociologiques ou psychologiques, les théories interprétatives en particulier.

Toutefois, les raisons qui sauvent ces théories sociologiques ou psychologiques sont aussi celles qui pourraient interdire à ces dernières de s'appliquer au domaine éthique : l'autonomie de l'éthique ne se justifie pas autrement que celle des différentes sciences humaines.

Autrement dit, ce sont les mêmes raisons qui pourraient, en principe, justifier certaines formes d'explication en sciences humaines et limiter leurs prétentions.

Mais il faut bien reconnaître, hélas, que nous n'en sommes pas encore arrivés au point où les physicalistes d'une part, les sociologues et les psychologues non réductionnistes de l'autre, accepteront de donner l'impression de vouloir limiter leurs prétentions. Les physicalistes n'ont certainement pas renoncé à l'ambition

d'éliminer les explications de type sociologique ou psychologique. Quant aux sociologues et aux psychologues, ils continuent, à quelques exceptions près, d'estimer qu'il vaudrait mieux se débarrasser, une bonne fois pour toutes, des élucubrations « éthiques ». J'aimerais pouvoir les convaincre que cette attitude n'est pas tout à fait cohérente.

PREMIÈRE PARTIE

LES CAUSES ET LES RAISONS

1. PRÉDICTIONS ET PLATITUDES

Si vous voulez vous faire une première idée de ce que signifie une explication par les raisons, pensez, tout simplement, au genre de réponse qu'il peut vous arriver de donner, lorsqu'on vous pose la question « Pourquoi ? ». On vous demande, par exemple, « Pourquoi avez-vous ouvert la fenêtre ? » et vous répondez : « Parce qu'il faisait chaud ». À première vue, c'est une « bonne » raison, même si elle est assez plate. On pourrait trouver plus savoureuse et plus honnête une réponse du genre : « Parce que je voulais contrarier un invité frileux que je déteste ».

Mais le problème philosophique resterait le même. Les réponses invoquent des raisons pour expliquer des actions. Est-il légitime de faire jouer à ces raisons, celles qui nous apparaissent immédiatement ou bien celles qui ne nous viennent à l'esprit qu'après un certain effort, le rôle de causes authentiques de nos actions, comme nous avons, semble-t-il, tendance à le faire naturellement ? Nous le faisons d'autant plus volontiers que nous saisissons plus clairement ce que signifie, en réalité, avoir une raison d'agir. Une raison d'agir, c'est une combinaison de désirs et de croyances. « J'ai ouvert la fenêtre parce qu'il faisait chaud » est une ellipse pour « J'ai ouvert la fenêtre parce que, d'une part, je désirais qu'il fasse moins chaud et, d'autre part, je croyais qu'il ferait moins chaud si j'ouvrais la fenêtre ». Habituellement, ce genre de précision est superflue, mais elle nous aide à comprendre pourquoi on a tendance à penser qu'une raison, consciente ou inconsciente, peut, pour ainsi dire, causer une action.

En accord avec ces platitudes de la psychologie ordinaire, certains philosophes estiment qu'il n'y a pas de différence de nature entre une explication telle que « J'ai ouvert la fenêtre parce qu'il faisait chaud » et d'autres telles que « La colonne de mercure s'est élevée dans le baromètre parce que la pression atmosphérique était plus forte ». Dans les deux cas, il s'agit d'une explication de type causal.

Les wittgensteiniens pensent que cette assimilation est un exemple typique de confusion conceptuelle, et qu'il faut être philosophe pour être sûr que ces deux « parce que », celui qui figure dans « J'ai ouvert la fenêtre parce qu'il faisait chaud » et celui qui se trouve dans « La colonne de mercure s'est élevée parce que la pression atmosphérique est plus forte » sont de même nature. Le premier « parce que » n'a rien de causal. En effet, une explication causale doit répondre à certaines exigences, qui ne peuvent pas être satisfaites, lorsqu'une raison, telle qu'« il faisait chaud » figure en tant qu'antécédent ou cause présumée.

1

D'une certaine façon, ce qu'on appelle, dans la littérature consacrée, le problème des causes et des raisons, c'est celui des deux « parce que »¹. Les choses se présentent donc plutôt bien,

1. À l'origine de la discussion sur les causes et les raisons, il y a un bref passage du *Cahier Bleu* de Wittgenstein (p. 14-15 de l'édition anglaise, parue chez Basil Blackwell en 1958, et p. 43-46 de la traduction de Guy Durand, Paris, Gallimard, 1965). Il y a, ensuite, ce que Vincent Descombes appelle joliment la série des petits livres rouges, publiés dans les années soixante aux éditions Routledge & Kegan Paul, et un petit livre « bleu » (ou plutôt verdâtre, dans les premières éditions) paru à Oxford. Le livre bleu c'est celui, fameux, d'Élizabeth Anscombe, *op. cit.*, 1957. Les livres rouges les plus directement liés à cette discussion sont ceux de Peter Winch, *The Idea of a Social Science* 1958 ; A.I. Melden, *Free Action* 1961 ; Anthony Kenny, *Action, Emotion and Will*, 1963, tous chez Routledge & Kegan Paul. Vincent Descombes a excellemment présenté le débat dans un numéro spécial de *La quinzaine littéraire* consacrée à la « philosophie anglo-saxonne », 644, 1- Avril, 1994. L'introduction de Pascal Engel à sa traduction de Donald Davidson, *Actions et événements*, Paris, PUF, 1993, est aussi très éclairante. Jean-Luc Petit s'est donné la peine d'essayer de bien faire comprendre les positions de Wittgenstein dans *L'action dans la philosophie analytique*, Paris, P.U.F., 1991, et Marc Neuberger également dans son introduction à *Théorie de l'action*, Bruxelles, Mardaga, 1991. Mais sur ces positions, on ne peut pas se passer des textes de Jacques Bouveresse : *Wittgenstein : la rime et la raison. Science, éthique et esthétique*, Paris, Minuit, 1973 ; et *Philosophie, mythologie et pseudo-science. Wittgenstein, lecteur de Freud*, Combas, L'éclat, 1991, dans lequel un chapitre entier est consacré à la discussion de la

c'est-à-dire assez simplement. Mais elles se compliquent un peu, du fait que les wittgensteiniens préfèrent ne pas donner de définitions des causes et des raisons, mais des moyens de les reconnaître (des critères d'identification, si on veut). À quoi reconnaît-on, d'après eux, qu'on est en présence d'une explication causale ? À trois choses, au moins.

Tout d'abord l'explication causale est l'aboutissement d'une procédure complexe. Vous observez, vous faites des hypothèses, vous confirmez vos hypothèses, vous répétez l'expérience en manipulant les facteurs qui vous semblent pertinents et, au bout du compte, vous devez réussir à montrer que la relation n'est pas une pure coïncidence et qu'elle est justifiée par une loi authentique. C'est ainsi qu'on procède, par exemple, pour savoir si c'est bien la pression atmosphérique ou un autre facteur qui élève la hauteur de la colonne de mercure dans un baromètre.

Ensuite, une explication causale met en relation des termes qu'on peut concevoir de manière indépendante. C'est bien pour cette raison que l'enquête causale commence par des hypothèses. Si on ne pouvait pas concevoir l'indépendance des termes, l'enquête n'aurait aucun sens. Il serait absurde de se demander s'il peut exister des montagnes sans vallées, alors qu'il n'est pas absurde de se demander s'il est possible de faire pousser des oranges au Groënland, même s'il est absurde d'essayer de le faire, étant donné ce que nous savons des oranges et du Groënland.

Enfin, les chaînes causales peuvent être indéfiniment allongées des deux côtés puisque toute cause peut être l'effet d'une autre cause et que tout effet peut être la cause d'un autre effet.

À présent, nous avons assez d'éléments pour comprendre pourquoi une explication par les raisons telle que « J'ai ouvert la fenêtre parce qu'il faisait chaud » ne pouvait pas être considérée comme une explication causale sérieuse par les wittgensteiniens. La réponse leur semblait évidente : c'est parce que les explica-

... dichotomie des causes et des raisons (p. 82-96). On peut lire, aussi, sur l'argument de Wittgenstein, le livre de Paul Johnston, *Wittgenstein and Moral Philosophy*, paru en 1989 chez l'éditeur des petits livres rouges. La tradition ne s'est pas perdue.

tions par les raisons ne présentent aucune des trois caractéristiques de l'explication causale : induction, indépendance logique, infinité de la chaîne explicative.

Premièrement, une raison n'est pas le genre de chose qu'on identifie au moyen d'observations, d'hypothèses, de confirmations. Revenons à notre exemple. On vous demande « Pourquoi avez-vous ouvert la fenêtre ? » Au lieu de répondre « parce qu'il faisait chaud », vous dites : « Je crois que c'est parce qu'il faisait chaud, mais je dois confirmer l'hypothèse, en répétant l'expérience, la vérifier progressivement, afin qu'au bout du compte je puisse certifier qu'elle est bien l'instance d'une loi disant "Chaque fois qu'il fait chaud, j'ouvre la fenêtre" ». C'est une réponse qui ne risque pas de passer inaperçue, mais qu'on ne prendra pas au sérieux, avec un peu de chance.

En fait, la connaissance que vous avez des raisons pour lesquelles vous faites telle ou telle chose est un peu comme la connaissance que vous avez de la position de votre propre corps. Pour savoir si vous croisez les jambes, vous n'avez pas besoin de les regarder. C'est une connaissance sans observation. Bref, si vous êtes en train d'observer, de faire des hypothèses, de rechercher des confirmations, c'est que vous n'êtes pas en train de chercher des raisons, ou que vous ne savez pas comment vous y prendre. Et même si la connaissance que vous avez de vos raisons d'agir n'est pas toujours aussi immédiate que celle que vous avez de la position de votre corps, elle appartient néanmoins au même genre de connaissance.

Deuxièmement, entre nos raisons d'agir et nos actions, il y a une sorte de connexion interne, de lien logique ou conceptuel. Cela veut dire, très approximativement, qu'il est difficile, ou impossible, de concevoir l'idée d'une raison d'agir sans que vienne aussitôt à l'esprit celle d'une action ou d'un effort dans une certaine direction. C'est aussi difficile ou impossible à concevoir qu'une montagne sans vallée. On peut le dire autrement en faisant observer que l'action (ou l'effort) n'est pas l'effet des croyances et désirs qui forment ma raison d'agir. C'est tout simplement une preuve, un critère d'identification de ces désirs et croyances.

Mon acte d'ouvrir la fenêtre est peut-être le seul moyen dont je dispose pour donner un contenu à mon désir et à ma croyance.

En réalité, il n'y a pas d'accès direct, parfaitement indépendant aux croyances et aux désirs : c'est toujours à travers une action ou un effort dans une certaine direction qu'ils s'expriment ou s'identifient. Mais si on ne peut pas concevoir de relation causale entre des termes qui ne sont pas indépendants, en un sens important quelconque, et si les raisons d'agir et l'action ne sont pas des choses indépendantes, il est absurde d'envisager une relation causale entre elles.

Enfin, la chaîne des raisons a un terme alors que la chaîne des causes n'en a pas. Pour bien comprendre cet argument, pensez à la différence entre la recherche d'une responsabilité et la recherche d'une cause ou d'un effet naturel. Si, en ouvrant la fenêtre pour rafraîchir la pièce, vous faites tomber un pot de fleurs sur la tête d'un passant, vous pourrez vous défendre en disant que vous n'aviez pas l'intention de le faire, même si cette défense risque de paraître suspecte aux yeux de certains psychologues qui sont toujours prêts à aller plus loin dans l'attribution de responsabilité. En réalité, vous n'êtes pas disposé à accepter toute la chaîne des conséquences de vos actes, mais seulement les conséquences que vous pouviez raisonnablement envisager ou celles que vous étiez préparé à accepter. Dans l'autre sens, vous êtes peut-être préparé à endosser la responsabilité de votre acte alors que vous auriez peut-être pu vous en décharger en accusant vos parents de vous avoir fait ou éduqué d'une façon telle que vous ne pouviez pas ne pas commettre ce genre de maladresse.

En résumé, il y a une fin dans la recherche de la responsabilité. On demande qui est l'auteur. On ne demande pas qui est l'auteur de l'auteur et ainsi de suite. D'autre part, on n'impute pas à l'auteur l'infinité des conséquences de son acte, mais une toute petite partie seulement de ces conséquences. Dans la recherche des causes et des conséquences naturelles, rien n'interdit, en principe, une quête infinie. Et comme la recherche des raisons ressemble à la recherche d'une responsabilité, elle n'a rien à voir, en principe, avec l'activité qui consiste à rechercher des causes.

Davidson a provoqué une certaine sensation en s'attaquant à ces trois arguments¹. Il a montré avec succès, semble-t-il, que les critères de distinction entre les causes et les raisons proposés par certains wittgensteiniens (mais pas par Wittgenstein lui-même, il faut le dire) n'étaient absolument pas suffisants ou décisifs. Puis il a essayé, avec un peu moins de succès peut-être, de montrer que mélanger les explications par les causes et les explications par les raisons, à la manière de Freud par exemple, ce n'était pas commettre un péché conceptuel épouvantable, mais se placer dans une situation épistémologique un peu inconfortable². En d'autres mots, Davidson a modifié le diagnostic. Il a substitué « problème épistémologique » à « confusion conceptuelle ».

Mais de nombreux philosophes, bien disposés pourtant, estiment que Davidson n'a pas franchement réussi à montrer comment ce problème pouvait être surmonté. C'est son échec en un certain sens. Mais commençons par les succès, c'est-à-dire par les arguments qui tendent à montrer que la dichotomie des causes et des raisons est assez floue, pour ne pas dire parfaitement injustifiée.

En premier lieu, il est faux de dire que la procédure qui nous sert à établir une relation causale est, par définition, inductive, qu'elle revient à construire des hypothèses, les soumettre à l'épreuve des faits, les confirmer progressivement, et qu'une relation ne peut être dite causale que si et seulement si on peut montrer qu'elle est une instance, un exemple d'une loi générale. Wittgenstein lui-même admettait qu'on pouvait parler d'identification directe, instinctive pour ainsi dire, de la causalité³. On suit le

1. Davidson D., « Actions, raisons et causes », (1963), dans Davidson D., *op. cit.*, 1993, p. 15-36. (Il existe aussi une traduction de Marc Neuberger dans son recueil, *op. cit.*, 1991, p. 61-78, sous le titre : « Actions, raisons d'agir et causes »).

2. Il faudrait citer la plupart des articles, rassemblés dans le recueil *Actions et événements*, ainsi que les trois essais, traduits par P. Engel et publiés ensemble aux Éditions de l'éclat sous le titre *Paradoxes de l'irrationalité* en 1991.

3. Cf. Bouveresse J., *op. cit.*, p. 83-84.

déplacement d'un objet tiré par un fil, on se brûle en posant sa main sur une plaque électrique incandescente. On n'a pas besoin de répéter l'expérience pour connaître la relation causale.

L'argument de Davidson est un peu plus complexe. D'après lui, le problème de l'induction ne se pose que lorsque nous voulons établir des lois causales. Mais la plupart du temps nous avons affaire à des relations causales singulières que nous avons souvent beaucoup de difficultés à justifier par de bonnes lois. On peut être sûr de l'existence d'une relation causale entre deux événements, par exemple l'impact du coup de poing et la fracture du nez, le heurt d'une pierre et le bris de la vitrine, sans être sûr de l'explication que nous donnons de la relation entre ces deux événements, qui dépend de la qualité de nos théories. C'est une distinction extrêmement importante pour Davidson. Elle détermine deux niveaux d'analyse : celui de la relation causale entre événements, qui est robuste, solide, et celui de l'explication causale qui est plus fragile, faillible, parce que formulée dans des langages théoriques concurrents, parfois incohérents en apparence, comme le langage des raisons qui sont des causes.

En fait, Davidson ne nous demande qu'une chose au plan ontologique : accordez-moi qu'il y a des événements. Acceptez que les collisions, les explosions, les banquets, sont des entités aussi respectables que les voitures, les volcans, les tables et les nappes blanches. Davidson n'exige qu'une caractérisation minimale de l'événement : il est cause ou effet, et ses propriétés dépendent du genre de description que nous voulons en donner. La relation entre événements est donc causale et indéterminée : nous pouvons la décrire comme nous le désirons, en termes physiques ou mentaux, c'est sans importance. L'explication causale est précisément le choix d'un certain type de description de la relation causale, encore indéterminée. C'est au niveau de l'explication causale seulement qu'on peut introduire des propriétés aussi déterminées que les propriétés mentales qui caractérisent les croyances et les désirs, dont la combinaison forme nos raisons d'agir. Ce n'est donc qu'au niveau de l'explication qu'il est légitime de parler de la causalité par les raisons. Mais cette explication ne peut que saisir approximativement ou partiellement (sous une certaine description) ce qui se passe au plan des événements. Ce qui se passe au plan des événements, nous pouvons,

en un certain sens, le saisir directement, sans explication. L'explication se superpose, pour ainsi dire, à la saisie directe, singulière des relations entre événements. C'est elle qui doit satisfaire aux règles habituelles de l'induction.

En résumé, cet ensemble d'arguments tend à montrer que du fait qu'une relation est saisie du premier coup, directement, sans induction, on ne peut nullement conclure que ce n'est pas une relation causale. Et on peut ajouter ce corollaire : du fait qu'on saisit en général la relation entre nos raisons d'agir et nos actions immédiatement, directement, on ne peut nullement conclure que ce n'est pas une relation causale.

En second lieu, il est faux de dire que l'explication causale n'a absolument rien à voir avec une attribution de responsabilité. C'est vrai de la relation causale singulière peut-être, mais l'explication causale est aussi sélective que l'explication par les raisons. C'est sous une certaine description seulement qu'une explication causale est valable. Je peux décrire un vase de Chine comme un cadeau de ma tante et comme un objet ayant une certaine masse. Supposons que je blesse ma tante en lui envoyant le vase de Chine sur le crâne. Je peux dire indifféremment que c'est le cadeau de ma tante ou un objet ayant une certaine masse qui lui a fendu le crâne, si je m'en tiens à la description de la relation causale. Mais je ne peux pas dire que c'est *parce que* c'est un cadeau de ma tante que le vase de Chine lui a fendu le crâne. C'est en tant que le vase a une certaine masse qu'il a pu causer cet effet brutal. Ce qui est pertinent dans l'explication causale, ce n'est pas la propriété d'être un cadeau de ma tante, mais d'avoir une certaine masse. C'est la raison pour laquelle on dit que l'explication causale est intentionnelle, puisqu'on ne peut pas substituer toutes les descriptions possibles de l'événement sans risquer d'annuler la valeur de l'explication (si on s'en tient, bien sûr, au « test » logico-linguistique courant — et contesté — de l'intentionnalité, illustré, généralement, au moyen de cet exemple dont on a peut-être trop abusé : si Cicéron n'est autre que Marcus Tullius, on devrait pouvoir dire, du fait que Cicéron est un bon orateur, il suit que Marcus Tullius est un bon orateur. Mais en préfixant ces phrases du verbe croire, la situation se modifie : elle devient « opaque ». Du fait que je crois que Cicéron est un bon orateur, il ne suit pas que je crois que Marcus Tullius est un bon orateur).

De tout ceci, on peut conclure que ce n'est certainement pas le caractère intentionnel d'une explication qui lui interdit d'être causale. À quoi on peut ajouter le corollaire suivant : du fait que l'explication par les raisons est intentionnelle (c'est ce qui autorise la comparaison avec une attribution de responsabilité) on ne peut pas conclure qu'elle ne peut pas être une explication causale digne de ce nom.

En troisième lieu, même s'il est incontestable que l'explication par les raisons présente des aspects normatifs ou des liens conceptuels, il n'en résulte pas qu'elle n'a absolument rien de causal. Dans les explications physiques, aussi, il y a des aspects normatifs, ceux qui relèvent des systèmes de mesure, par exemple, mais cela ne nous interdit pas de penser qu'elles peuvent être des explications causales authentiques.

Pour qu'une explication n'ait rien de causal, il ne suffit pas qu'elle ait des aspects logiques ou conceptuels, il faut aussi qu'elle ne tienne aucun compte de ce qui peut se passer à un autre niveau d'analyse, et que les liens entre ces niveaux soient inconcevables.

3

En conclusion, étant donné qu'une explication peut être causale, même si elle est non inductive, dépourvue de couverture légale, intentionnelle, partiellement normative, rien ne nous interdit d'affirmer que l'explication par les raisons est causale, en dépit du fait, bien établi par les wittgensteiniens, qu'une explication par les raisons est non inductive, dépourvue de couverture légale, intentionnelle et partiellement normative. Bref, si les wittgensteiniens ont eu raison à propos des caractéristiques de l'explication par les raisons, ils ont eu tort d'en conclure qu'elle ne pouvait pas être causale.

Si on est assez attentif, on s'aperçoit que Davidson substitue à la métaphore de la différence des jeux de langage, la métaphore des niveaux d'analyse. Au premier étage, si on peut dire, il y a la relation causale qui vaut entre les événements indéterminés. Au second étage, l'explication causale formulée dans des langages théoriques concurrents parfois incohérents.

Pour le wittgensteinien, si on mélange les jeux de langage, on se perd. Pour Davidson, si on mélange les niveaux d'analyse, on se perd aussi. Or Davidson estime que l'erreur des wittgensteiniens c'est précisément d'avoir confondu les niveaux d'analyse pour éviter une prétendue confusion dans les jeux de langage. Pour Davidson, le problème des raisons qu'on prend pour des causes, s'il y en a un, ne peut se poser qu'au deuxième niveau, celui de l'explication. Et s'il n'est pas tout à fait insoluble, c'est, entre autres choses, parce qu'on peut, à tout moment de l'analyse, revenir au premier niveau, celui de la relation causale authentique entre événements. Je ne vais pas rentrer ici dans le détail de ces analyses extraordinairement complexes qui font appel à toutes sortes de concepts dont la valeur est intensément discutée, et qui ont, aujourd'hui encore, un petit air de science-fiction. C'est le cas, en particulier, de la « survenance », concept dont il est fait grand usage dans cette justification de l'analyse en niveaux, et qui devrait, en principe, éloigner la menace de l'inertie causale ou de l'épiphénoménalisme du niveau supérieur, (le niveau « mental ») ou celui de l'explication par les raisons. Mais il est possible, tout de même, de dire quelque chose de cette difficulté épistémologique (substituée à la confusion conceptuelle) sans pénétrer dans ce marais ontologique.

4

On l'a vu, rien n'interdit qu'une explication par les raisons soit causale, même si elle est intentionnelle, normative, non inductive, et dépourvue de « couverture » légale. Toutefois, dire d'une explication d'un certain genre qu'elle est dépourvue de couverture légale ou qu'elle ne peut pas être justifiée par une loi (ayant l'allure d'un conditionnel universel), c'est donner à cette explication un statut purement rétrospectif. Car on ne peut parler sérieusement de prédiction, s'il n'y a pas de loi pour la justifier. Cette restriction devrait s'appliquer aux explications causales des comportements par les raisons. En effet, si l'explication causale par les raisons est, par nature, dépourvue de couverture légale, comme le prétend Davidson, cela veut dire que nous ne pouvons pas espérer prédire sérieusement la conduite de quelqu'un à partir de ses désirs et de ses croyances. On peut tout

juste l'expliquer rétrospectivement, ce qui n'interdit pas que cette explication soit causale. Dans la vie quotidienne, ce n'est peut-être pas un déficit considérable. Apparemment, on se débrouille très bien au moyen d'anticipations vagues, inexprimées. On a du mal à voir ce que des prédictions sérieuses, précises, pourraient ajouter : elles risqueraient plutôt de donner un air ridicule aux relations humaines, un air qui deviendrait très vite le seul intérêt que ces relations pourraient encore présenter. Mais, l'impossibilité de confectionner des lois et celle, corrélative, de prédire, interdisent d'aligner les sciences humaines sur les sciences de la nature.

Bref, il n'y a rien d'extravagant ou de confus dans l'explication causale de l'action par les raisons. Il n'y a qu'un déficit épistémologique : l'absence de capacités de prédiction, dérivées de l'impossibilité de confectionner des lois, lorsqu'un des termes de l'explication est mental, comme c'est le cas dans l'explication par les raisons. Mais pourquoi l'intervention de termes mentaux (les « raisons » étant des complexes de choses mentales : désirs et croyances) dans l'explication interdit-elle la confection de lois ?

Davidson avance plusieurs arguments. Je vais me contenter d'examiner celui dont la portée est la plus large et touche l'ensemble des théories en sciences humaines. Davidson semble rejeter les prétentions des sciences humaines à être des sciences sous le prétexte qu'elles ne proposent pas et ne pourront jamais proposer des lois strictes. Fodor et d'autres philosophes rejettent cet argument¹. Ils estiment que toutes les sciences spécialisées, géologie, biologie, etc., sont exactement dans la même situation, ce qui ne nous empêche pas de les considérer comme des sciences tout à fait respectables. En fait, Davidson justifie son argument de deux façons différentes. D'une part, il montre que toutes les prétendues lois psychologiques introduisent des clauses de sau-

1. Fodor J., *Psychosemantics*, Cambridge, Mass., MIT Press, p. 4-5, et « Everything Else Being Equal ; Hedged Laws and Psychological Explanation », *Mind*, 1, 1991, p. 21-34. Pour une discussion approfondie voir Stephen Schiffer, « Ceteris Paribus Laws », *Mind*, 1, 1991, p. 1-17 ; Pierre Jacob, « Patterns, causalité mentale et lois intentionnelles », *Les études philosophiques*, 3, 1992, p. 391-413 ; Élisabeth Pacherie, *Naturaliser l'intentionnalité*, Paris, PUF, 1993, p. 65-102.

vegarde exprimées dans un vocabulaire non psychologique. D'autre part, il montre que les prétendues lois psychologiques sont des sortes de truismes, de vérités *a priori*. Il me semble que la première justification de Davidson tombe sous le coup de la critique de Fodor, mais pas la seconde. Personnellement, je préfère la seconde bien qu'il me semble qu'elle ait aussi ses limites. Je vais essayer de donner un peu de chair à cette discussion au moyen d'un exemple emprunté à la psychologie sociale.

5

Il existe, en psychologie sociale, un ensemble d'expériences qui tendent à montrer que notre propension à aider des personnes en détresse varie avec la quantité de témoins qui observent la scène¹. Cette variation est paradoxale : plus il y a de témoins et moins on aide ; et réciproquement moins il y a de témoins et plus on aide. Étant donné que cette relation semble satisfaire aux critères de légalité minimaux (c'est une généralisation non accidentelle qui peut subir avec succès des tests contrefactuels du genre : s'il n'y avait pas eu de témoins, il aurait aidé la victime), je n'hésiterai pas à l'appeler « loi », suivant en cela l'exemple de Fodor, qui dans sa discussion donne le nom de « loi » à une relation qui a exactement la même portée, comme nous allons le voir.

À première vue, ce que dit cette « loi » psychologique contredit nos idées habituelles sur la question. Pensez à ce qui se passe dans les ventes de charité ou dans les festivals philanthropiques à la télévision. Vous en conclurez : plus il y a de témoins et plus on aide. Mais le caractère paradoxal de la « loi », plus il y a de témoins, moins on aide, la rend d'autant plus intéressante.

Apparemment, elle est homogène, entièrement formulée en termes de comportements observables. Il n'y a aucune référence à des croyances, des désirs, des pensées et autres choses qu'on ne peut ni voir, ni entendre, ni toucher, ni même appréhender indirectement et mesurer avec une précision suffisante.

1. Nisbett R., et Wilson T., « On Telling More than We Can Know : Verbal Reports on Mental Processes », *Psychological Review*, 84, 1977, p. 231-259 ; cf. aussi la discussion de Paul Johnston, dans *op. cit.*, 1989, p. 46-49.

Par ailleurs, il n'y a aucune raison de supposer que les observateurs se sont moqués du monde, c'est-à-dire qu'ils n'ont pas défini de façon claire et opérationnelle certains termes problématiques, tel que le verbe « aider » (afin d'éviter la confusion avec la charité télévisée, entre autres).

Enfin, rien n'interdit d'envisager la possibilité de prédire à partir de cette loi et de la confirmer en multipliant les expériences. Bref, c'est de la vraie science. Enfin, presque. Supposons en effet que l'un ou l'autre des sujets d'expérience ait cru seulement qu'il y avait des témoins alors qu'il n'y en avait pas, et qu'il se soit abstenu d'agir pour cette raison. Mais l'observateur a choisi de s'en tenir au comportement extérieur, public, accessible. Il s'est donné pour règle de ne rien demander aux sujets d'expérience ; il n'a que faire de leurs prétendues croyances. Il observe, comme s'il avait affaire à du papier de tournesol. Il risque d'être embarrassé. Car il devra tenir compte d'une exception manifeste à sa « loi », puisque quelqu'un n'aura pas aidé une personne en détresse alors qu'objectivement il n'y avait pas de témoins. S'il voulait sauver sa « loi », il lui suffirait d'introduire des « clauses de sauvegarde », des réserves, disant par exemple « à moins que les sujets d'expérience se trompent », « à moins qu'ils aient des croyances fausses », « à moins qu'ils aient des hallucinations ». Mais notre observateur veut s'en tenir au comportement public. C'est un behavioriste comme on dit. Il n'est, en principe, absolument pas équipé théoriquement pour parler de ces choses, puisqu'il n'a rien à dire, en tant que scientifique, sur les croyances, même s'il en parle tout le temps après ses heures de travail.

Supposez qu'un mentaliste, c'est-à-dire quelqu'un qui accepte dans ses explications ou dans son ontologie de faire intervenir des phénomènes mentaux tels que les croyances et les désirs, prenne le relais du behavioriste, désespéré par les limites de sa théorie. Il risque de lui arriver la même chose : il va nous parler de croyances, à moins que le cerveau se détraque. Mais que sait-il du cerveau ? Dans le cadre de sa théorie, rien, car il n'est pas neurophysiologiste. Cela ne l'empêchera peut-être pas de parler, mais il dira probablement n'importe quoi ou il utilisera des clichés destinés au grand public dont il fait partie, pour le neurophysiologiste. D'autre part, on ne voit pas comment il pourra éviter de parler de comportements observables, gestes, mots, parce

qu'il est tout de même difficile d'identifier des croyances sans se servir de ce genre de témoignages objectifs. Mais que sait-il lui, le mentaliste, de la structure formelle du langage, des méthodes rigoureuses d'observation du comportement ?

D'après Davidson, il n'y a que dans les disciplines fondamentales, la physique par exemple, qu'on peut formuler des lois sans outrepasser son domaine de compétence si on peut dire, c'est-à-dire sans introduire subrepticement des termes parfaitement étrangers à la discipline. Une loi physique qui introduirait des croyances ne serait peut-être pas entièrement absurde, mais elle ne relèverait pas entièrement de la compétence du physicien, à moins que les croyances ne soient elles-mêmes réduites à des états physiques. Mais, heureusement pour le physicien, il peut formuler ses lois sans avoir à faire intervenir directement ou dans des clauses de sauvegarde des phénomènes sur lesquels il n'a pas de prise (croyances ou désirs par exemple).

Le psychologue n'a pas cette chance. Il ne peut pas formuler des lois du mental sans faire intervenir directement ou dans des clauses de sauvegarde des concepts qui relèvent d'un autre domaine. En réalité, il n'y a pas de véritable autonomie de l'explication psychologique. C'est même, en fait, l'ensemble des sciences humaines qui semble limité par ce genre de contrainte et c'est pourquoi leur situation paraît désespérée. Elle le paraît d'autant plus qu'il est absurde d'envisager la possibilité de se passer des sciences humaines. On est donc condamné à vivre longtemps encore avec ce genre de sciences bancales.

6

De nombreux philosophes n'acceptent pas ce diagnostic pessimiste. Certains ont répondu à Davidson que la situation des sciences humaines n'était absolument pas exceptionnelle. En réalité, dans toutes les sciences spécialisées, géologie, biologie, etc., on a recours au vocabulaire d'autres sciences ou de sciences plus générales : physique, chimie, pour déterminer les conditions d'application des « lois ». Dans ces sciences spécialisées, on formule fréquemment des lois approximatives qui ne valent qu'à condition d'être accompagnées de clauses de sauvegarde, de propositions auxiliaires, de compléments conditionnels, tout ce

qu'on résume au moyen de l'expression « toutes choses étant égales par ailleurs ». Fodor donne l'exemple de ce qu'il appelle une « loi » géologique : « Les méandres des rivières érodent la paroi externe de leurs berges ». On ne peut pas dire qu'il s'agisse d'une « loi » stricte ou d'une généralisation sans exceptions. Fodor estime que cette « loi » n'est vraie qu'accompagnée de compléments conditionnels, de clauses de sauvegarde telles que : si le climat ne change pas au point de faire geler les rivières, si personne ne s'avise de construire une digue, si le monde n'est pas détruit, etc. Et, bien entendu, aucune de ces clauses de sauvegarde n'est formulée dans le vocabulaire de la géologie au sens strict. Bref, selon Fodor, la psychologie n'est pas plus mal lotie que la géologie ou la biologie. Elle n'a rien d'exceptionnel, ni en bien, ni en mal. Mais Fodor a tort, comme d'habitude. Il y a tout de même quelque chose d'exceptionnel dans les prétendues « lois » du comportement.

Prenons la soi-disant prédiction « Si Édouard veut revoir sa maîtresse et qu'il croit qu'elle est à Lyon, il ira à Lyon, s'il ne change pas d'avis, si les trains fonctionnent, si son cerveau ne se détraque pas, etc. » Nous avons introduit des réserves, ce qu'on appelle des clauses de sauvegarde comme dans les autres prédictions empiriques. Mais, en réalité, la difficulté porte sur le premier conditionnel. Quand je dis « Si Édouard veut revoir sa maîtresse et qu'il croit qu'elle est à Lyon, il ira à Lyon », je ne fais pas une prédiction à proprement parler. Aller à Lyon n'est pas la conséquence de son désir et de sa croyance, c'est plutôt une preuve du fait qu'il avait ce désir et cette croyance, ou plus exactement du fait que son comportement est rationnel, cohérent.

Mais la cohérence ou la rationalité globale des croyances, des désirs et des comportements n'est pas une hypothèse qui pourrait être démentie par l'expérience ! C'est un principe qui est au centre de tout notre système de croyances, un principe auquel nous ne pouvons pas renoncer et devant lequel l'expérience doit, pour ainsi dire, se plier : les cas de mentalité « pré-logique » ou « pré-rationnelle », sont des cas d'incompétence de l'ethnographe. Le principe de rationalité ne peut pas être démenti, l'honnêteté ou le talent de l'ethnographe, si. Bref, la rationalité (ou la cohérence) des comportements n'est pas, à proprement parler, un fait d'observation, c'est une norme qui l'oriente. Les pseudo-prédic-

tions qui ressortissent de ce principe de rationalité (telle que le conditionnel : « Si Édouard veut revoir sa maîtresse » etc.) ne peuvent être ni améliorées par la réduction des clauses de sauvegarde, ni démenties par l'expérience.

Il n'y a rien de réfutable dans « Si Édouard veut revoir sa maîtresse et qu'il croit qu'elle est à Lyon, il ira à Lyon ».

En fait, des propositions dérivées de ce genre de conditionnel ne peuvent pas être fausses. Il en va tout autrement pour les conditionnels du genre de ceux qu'évoque Fodor. La relation entre le tracé d'une rivière et le processus d'érosion est une authentique découverte, un fait d'observation. On ne peut pas la connaître d'avance et il n'est pas absurde de supposer qu'elle est fausse.

Il est difficile de croire que les clauses de sauvegarde (toutes choses étant égales par ailleurs) jouent le même rôle en complément de ces énoncés, qui peuvent être faux, et en complément d'énoncés qui ne peuvent pas l'être.

La difficulté épistémologique tout à fait propre aux sciences humaines est donc assez bien identifiée. C'est une difficulté qui fait de ces sciences un mélange assez curieux de platitudes ou de généralités irréfutables et d'observations assez intéressantes mais chaotiques de faits singuliers.

Davidson semble osciller entre deux attitudes : faire des conditionnels du genre « Si Édouard veut revoir sa maîtresse et qu'il croit qu'elle est à Lyon, il ira à Lyon », ou bien une sorte de loi approximative ou bien une sorte de vérité conceptuelle. Il me paraît évident que la seconde attitude est la seule qui soit raisonnable. D'autre part, je crois qu'il faudrait aussi envisager la possibilité de traiter ces conditionnels comme des sortes de règles de conduite. De ce point de vue, une analyse de phrases telles que « Si Édouard a promis d'aller à Lyon, il ira à Lyon », serait infiniment plus fructueuse. Bref, il faudrait distinguer, me semble-t-il, le couple loi approximative — clauses de sauvegarde, des truismes et du couple règles — manquements à la règle. Dans les sciences humaines, c'est à des truismes et à des couples règles-manquements à la règle qu'on a affaire principalement. Je reviendrai sur ce sujet dans l'examen des trois aspects de l'efficacité rituelle.

Mais je ne voudrais surtout pas conclure ce chapitre sur une note d'optimisme philosophique, en laissant entendre que le

changement d'orientation de Davidson, le passage de la confusion conceptuelle à la difficulté épistémologique représenterait un authentique « progrès », ou que les arguments de Davidson auraient définitivement balayé ceux de Wittgenstein (ou plutôt des wittgensteiniens).

7

Les philosophes sont nombreux à penser que la difficulté épistémologique identifiée par Davidson est parfaitement insurmontable. Il leur arrive aussi d'ajouter que c'est peut-être bien parce qu'il existe une authentique confusion conceptuelle dans l'explication causale par les raisons, qu'elle pose tant de problèmes.

Quoi qu'il en soit, il y a une chose qui n'a jamais cessé de réunir partisans et adversaires de la dichotomie des causes et des raisons : un profond scepticisme à l'égard de la possibilité d'aligner les sciences de l'homme sur les sciences de la nature. Dans les premières, ce qui nous intéresse, ou ce que nous ne pouvons pas réduire, bref ce que nous ne voulons pas ou ne pouvons pas éliminer, c'est précisément ce qui ne peut pas être enrégimenté dans une explication du genre de celles qu'on trouve dans les sciences de la nature par les causes et par les lois.

Pourtant, nous nous obstinons à vouloir sauver la causalité par les raisons. C'est que nous restons toujours plus ou moins prisonniers de la métaphore du voyage.

2. SATISFACTION ET RÉFLEXION

Parmi les métaphores que les philosophes utilisent, plus ou moins explicitement, dans leurs descriptions de l'action humaine, c'est celle du voyage qui est probablement la plus populaire.

Je voudrais essayer de montrer que cette métaphore, dont l'utilité et la fécondité sont difficilement contestables, est aussi à l'origine d'innombrables difficultés. Elle est, pour ainsi dire, la source de la plupart de nos perplexités, quand nous réfléchissons sur ce qu'agir veut dire ou sur ce qui nous permet de distinguer ce que nous faisons et ce qui nous arrive. En réalité, à travers la critique du modèle de la satisfaction, ce sont les limites de la métaphore du voyage que j'essaierai de mettre en évidence.

L'expression métaphore du voyage n'a pas été utilisée, à ma connaissance, pour qualifier l'image de l'action humaine que se font certains philosophes, certains sociologues, certains psychologues. Mais il me semble qu'elle est assez appropriée pour désigner l'idée qui les oriente lorsqu'ils présentent l'action comme une sorte de parcours à obstacles. Pour agir au sens plein, il faudrait, en quelque sorte, suivre un itinéraire prévu d'avance et franchir, l'une après l'autre, les étapes suivantes : désirer — délibérer — choisir — commencer d'exécuter — accomplir sans renoncer.

C'est un schéma qu'on a pris l'habitude d'attribuer à Aristote, en dépit du fait qu'il n'y a jamais eu d'unanimité sur le point de savoir si le Philosophe avait réellement distingué des étapes, ou lesquelles et en quel ordre. Ces incertitudes n'ont pas empêché Tricot, par exemple, d'employer l'expression « étapes de l'acte moral », pour présenter la description qu'Aristote avait proposée de l'action volontaire dans l'Éthique à Nicomaque¹.

1. Aristote, *L'Éthique à Nicomaque*, trad. J. Tricot, Paris, Vrin, 1983, note p. 137.

Certains philosophes ont souligné le caractère métaphorique des descriptions de l'action en tant que succession d'étapes, mais sans employer l'expression métaphore du voyage. Je pense à Austin, en particulier¹.

En tout cas, on pourrait dire que ce qui donne son tracé spécifique à ce parcours à obstacles, cette succession d'étapes, c'est que son point de départ est intérieur et son lieu d'arrivée extérieur. En effet, les premières étapes se déroulent en nous (puisque c'est là que se forment nos désirs, nos croyances, c'est là que nous délibérons et choisissons) et la dernière est pour ainsi dire, hors de nous, puisqu'elle s'exprime dans la modification d'un certain état de choses dans le monde ou dans l'esprit d'autrui. Ou disons que tel doit être le cas lorsqu'on prend pour paradigme de l'action, l'action transitive, fabricatrice ou stratégique, et non l'action immanente, ou expressive. J'ai dit qu'il s'agissait d'une métaphore, car, aussitôt qu'on examine ce schéma de plus près, aussitôt qu'on essaie de prendre l'image à la lettre, son caractère lacunaire, sélectif, imprécis ne peut pas manquer de sauter aux yeux.

Comment faisons-nous, par exemple, pour fixer les limites entre l'action et ce qui la précède, ou entre l'action et ce qui la suit ? À partir de quel moment sommes-nous *dans* l'action ? Quand commence-t-elle, où s'achève-t-elle ? Les intentions, décisions, choix, actes de volonté, contractions musculaires, impulsions nerveuses sont-ils des parties de l'action ou des *préliminaires* de l'action seulement ? Si certaines étapes de l'action sont elles-mêmes des actions, ce qui les suit ne doit-il pas être considéré comme une conséquence de l'action ? Bref, comment faisons-nous pour distinguer l'action, en tant que suite de séquences ordonnées, et ce qui, tout en étant lié à l'action d'une certaine façon, doit être rangé plutôt dans la classe des antécédents ou des conséquences de l'action ?² À ces problèmes qui restent en suspens, en dépit de l'énorme quantité d'énergie intellectuelle qui a été dépensée pour les clarifier ou les résoudre, s'ajoute celui

1. Austin J.L., *Écrits philosophiques* (1970), trad. L. Aubert et A.L. Hacker, Paris, Le Seuil, 1994, p. 156-157 et 241-243.

2. A. Oldenqvist, « Choosing, Deciding and Doing », *The Encyclopedia of Philosophy*, Edwards P., éd., New York, The Free Press, McMillan, 1967.

de l'enchaînement des étapes de l'action. Même si on admet que la métaphore du voyage doit être sauvée, que l'objet de toute théorie de l'action consiste en définitive à essayer de justifier rationnellement cette métaphore, on peut se poser des questions sur ce qui explique le passage des premières étapes aux suivantes. Nous avons à notre disposition deux théories pour expliquer cet enchaînement.

La première est internaliste. Elle postule l'existence d'une sorte de lien logique ou de connexion interne entre le vouloir-faire et son actualisation, si ce vouloir-faire est éclairé ou sincère. Il est assez difficile de comprendre exactement ce que signifie l'idée du lien logique ou de la connexion interne. S'agit-il d'une norme présupposée dans toute description d'action humaine ? S'agit-il d'une sorte de contrainte subjective, similaire à celle qui nous retient de nous contredire ? Les deux, probablement.

La seconde explication de l'enchaînement est externaliste. Entre les raisons d'agir et l'action, il n'y aurait pas de liaison nécessaire mais une liaison causale ordinaire, probable ou contingente. Et, c'est ce qui expliquerait pourquoi il nous arrive, parfois, de ne pas faire ce que nous avons jugé bon de faire. L'antécédent mental serait suivi de son conséquent physique dans la plupart des cas, mais pas dans tous les cas.

On a pris l'habitude d'attribuer à Socrate la paternité de la thèse internaliste. C'est elle qui serait censée s'exprimer dans le fameux paradoxe : « Nul n'est méchant volontairement ». Et c'est souvent par le biais de la critique de ce paradoxe qu'on aboutit au rejet de la thèse internaliste. Les internalistes n'ont jamais réussi à expliquer le fait indéniable qu'il nous arrive de ne pas faire ce que nous avons jugé bon de faire, même lorsqu'aucun autre désir plus puissant ne s'opposait au choix initial et même lorsque rien ne s'opposait à l'exécution de ce choix. De leur côté, les externalistes n'ont pas donné d'explication convaincante à la distribution des cas. Pourquoi nous arrive-t-il parfois de faire ce que nous avons jugé bon de faire et parfois pas ?

Pour remédier à toutes ces difficultés et pour expliquer l'enchaînement correct des étapes de l'action, il a paru utile de postuler l'existence de forces d'entraînement ou de liaison. C'est à ces forces (ces entités intermédiaires) qu'on a souvent choisi de donner le nom générique de « volonté », et plusieurs définitions

spécifiques différentes, ce qu'on peut comprendre aisément en observant que chacune est relative à une étape différente du parcours.

À l'étape du désir, la volonté, c'est l'impulsion initiale plus ou moins raisonnée, celle qui nous oriente vers des biens dont tout le monde reconnaît l'importance. À l'étape de la délibération, la volonté, c'est cette faculté de se déterminer, de choisir même lorsque les options semblent équivalentes ou le meilleur parti difficile à estimer. À l'étape de l'action, la volonté, c'est d'abord l'effort pour essayer d'obtenir ce qui a été choisi ou décidé. Et puis, c'est une certaine façon d'aller jusqu'au bout de sa tâche, fût-elle pénible, sans se laisser distraire et sans se détourner au premier obstacle. Bref, c'est une certaine manière « tenace », « persévérante » de se comporter. L'adverbe « volontairement » peut alors se substituer, raisonnablement, au substantif « volonté ». Le terme « volonté » est suffisamment élastique pour supporter toutes ces métamorphoses — et d'autres encore — ce qui agace beaucoup certains philosophes qui se contenteraient bien de préserver le dernier sens adverbial (ou comportemental) en se débarrassant des usages « substantifs » ou « mentaux » de la volonté¹. Quoi qu'il en soit, l'idée de volonté donne incontestablement les moyens de résoudre l'énigme des actions qui ne suivent pas les décisions. Le partisan des théories volitionnelles dira : si vous voulez savoir pourquoi il arrive qu'une personne ne fasse pas ce qu'elle jugeait bon de faire, examinez l'état de sa volonté. Ce qui explique la rupture du lien entre les étapes internes et externes de l'action, c'est la faiblesse de la volonté. Et, de même qu'une volonté forte prend un visage différent à chacune des étapes du parcours de l'action, une volonté faible présente des traits différents, selon l'étape où elle est diagnostiquée. Lorsque la défaillance a lieu à l'étape du désir, la faiblesse de la volonté, c'est l'absence d'impulsion ou d'appétit pour les biens dont l'agent reconnaît l'importance. Lorsque la défaillance se produit à l'étape de la délibération, la faiblesse de la volonté devient difficulté à conclure même lorsque le meilleur parti est

1. Je fais allusion, bien sûr, à Gilbert Ryle et à tous ceux qui le suivent. Ryle G., *La notion d'esprit* (1949), trad. S. Stern-Gillet, Paris, Payot, 1978, p. 67-72.

évident. Enfin, lorsque la panne a lieu à l'étape de l'exécution, la faiblesse de la volonté, c'est l'arrêt à l'endroit du passage à l'acte, l'incapacité à commencer à mettre en œuvre ce qui a été choisi ou décidé. Ou, un peu loin dans cette même étape (si on a accompli le prodige du passage à l'acte), c'est l'absence de ténacité, de persévérance, la tendance à abandonner en chemin une tâche dont l'agent a clairement reconnu l'importance et qui est manifestement à sa portée.

À cette longue liste de déficiences bien connues (et magnifiquement illustrées dans la littérature par ces figures emblématiques que sont Oblomov, Adolphe, Des Grieux), on peut ajouter une panne beaucoup plus générale et d'un ordre assez différent, celle que connaîtraient plutôt Achille ou Médée : c'est l'impatience ou l'impulsivité, c'est-à-dire la tendance à se précipiter dans l'action sans avoir délibéré. En dépit de leurs aspects « énergiques », les actions de ce type appartiennent cependant au genre « faiblesse de la volonté », car l'agent n'a pas maîtrisé son impulsion initiale (il y a donc une défaillance à l'étape du désir) ; il a négligé la délibération (il y a donc une déficience à la deuxième étape du voyage). Enfin, pour ces deux raisons, l'énergie dont il fait preuve éventuellement dans l'action n'est pas ténacité, endurance, persévérance, mais opiniâtreté ou obstination (il y a donc une troisième déviation, relativement à l'itinéraire normal ou idéal).

Le fait qu'on puisse considérer que certaines actions impulsives ou impatientes résultent de la faiblesse de la volonté montre bien que la volonté n'est pas conçue, en général, comme une sorte d'énergie brutale. La volonté en tant que poussée interne devrait être distinguée du désir, des élans aveugles et autres pulsions animales. C'est du moins ce qu'il conviendrait de faire (et ce qui n'est pas toujours fait comme on le constate aisément). Le personnage d'Hamlet est comme l'allégorie de toutes ces défaillances (asthénie, indécision, absence de persévérance, mais aussi impulsivité, impatience). Et il n'est pas très étonnant qu'Hamlet reste une source inépuisable de réflexions pour ceux qui s'intéressent à la faiblesse de la volonté, en tant que cause de nos actions déplorables.

Cependant, même si on admet que la volonté peut rendre tous ces excellents services, on peut tout de même lui reprocher d'être une entité assez mystérieuse ou, lorsqu'on en reste sobrement au

plan sémantique, d'être une notion confuse¹. La volonté est sûrement mystérieuse lorsqu'on dit d'elle qu'elle est une sorte d'entité spirituelle capable néanmoins d'agir sur des corps (pensez au magicien courbant des fourchettes à distance ou manipulant son audience sans quitter la scène). Mais la volonté peut nous embarrasser, même lorsque nous souhaitons lui faire jouer un rôle infiniment plus modeste. C'est ce qui arrive lorsqu'on introduit la notion de volonté pour décrire cette entité supplémentaire susceptible d'expliquer les pouvoirs moteurs du désir, les processus de conclusion d'une délibération, le passage à l'acte et la conduite de l'action de l'initiative à son terme. À chaque étape et entre chacune d'elles, il faudrait vouloir : vouloir son désir, vouloir son choix, vouloir passer à l'acte, puis vouloir l'action. Or, en introduisant ce genre d'entité supplémentaire, on s'expose ou bien à la régression infinie, ou bien à la contradiction. Pourquoi ? S'il faut vouloir son désir pour qu'il soit « moteur », vouloir son choix pour qu'il entraîne le passage à l'acte, vouloir l'action pour la mener à son terme, ne faut-il pas également vouloir tous ces vouloirs et vouloir ce vouloir de ces vouloirs et ainsi de suite à l'infini ? Mais on peut objecter que cette régression est exclue. Car de même que nous n'admettons pas l'idée qu'il soit possible de s'obliger à aimer, de se forcer à croire ou à désirer, nous rejetons l'idée qu'on puisse vouloir vouloir. Cependant, si l'idée d'un vouloir vouloir est exclue, la volonté devient une notion contradictoire. La volonté s'imposerait à nous, en quelque sorte, puisque nous ne pourrions que l'accueillir, et non la vouloir. Et ceci suffirait à rendre involontaire la plus volontaire, la plus souveraine, de nos actions.

On ne trouvera pas d'issue en réservant l'adjectif « volontaire » pour toutes ces actions qu'on peut commander à l'impératif (« Lève-toi et marche », « Ramasse les feuilles mortes », « Dis quelque

1. Ceux qui s'intéressent à ces problèmes, dits sémantiques, disposent désormais de la somme de Brian O'Shaughnessy, *The Will, A Dual Aspect Theory*, Tomes 1 et 2, Cambridge University Press, 1980. Il existe une traduction française du commentaire de H. D. Lewis, « L'esprit et le corps selon O'Shaughnessy », *Archives de philosophie*, 49, 1986, p. 75-97. On peut avoir l'impression que le bref et excellent article de Vesey en dit autant sur la nature des problèmes que pose l'idée de volonté, G.N.A. Vesey, « Volition », *Philosophy*, 36, 138, 1961, p. 352-365.

chose »). Le problème de l'anomalie conceptuelle de la volonté ne peut pas être résolu par ce moyen, car les composants essentiels de l'action volontaire seront dits involontaires selon ce critère. On ne peut pas commander « désire ! », « délibère ! », « forme des croyances relatives aux moyens de satisfaire tes désirs ! » etc.

Dans ces conditions, il n'est pas difficile de comprendre pourquoi certains philosophes, désireux de sauver la métaphore du voyage, l'image de l'action en tant que succession d'étapes, aient préféré se passer de la volonté. En un certain sens, ils ne prenaient guère de risques. Depuis Hume, au moins, les philosophes savent comment on peut respecter un schéma causal sans faire intervenir des forces liantes au statut douteux. Il suffit, semble-t-il, de donner une signification extrêmement modeste à l'idée de causalité. Aux forces efficientes, pouvoirs occultes, influences subtiles, on substitue l'idée d'une relation de consécution entre deux événements, l'apparition du premier (la cause, l'antécédent) permettant de prédire, avec des chances de réussite qu'il est possible d'estimer à l'avance, l'apparition du second (l'effet, le conséquent). Tout ce qu'on exige de ce schéma c'est qu'il donne les moyens de prédire correctement l'apparition des phénomènes. Quant aux hypothèses sur les « forces efficientes » ou même sur les mécanismes internes, elles sont, comme le disait Russell, des résidus d'une époque révolue¹.

Le schéma dit humien a été transféré à la théorie de l'action. C'est celui que j'appelle *modèle de la satisfaction*. C'est un schéma causal qui se passe d'entités intermédiaires. De nombreux arguments ont été dirigés contre ce schéma. J'insisterai surtout sur le fameux argument des chaînes causales déviantes. Il dit que le fait qu'un agent ait pris l'initiative de réaliser un certain état de choses, associé au fait que cet état de choses s'est réalisé, ne garantit pas que c'est bien l'agent qui l'a accompli. En termes plus généraux, l'argument des chaînes causales déviantes dit qu'on ne peut pas se contenter, pour décrire l'action humaine, d'établir une relation de correspondance entre un antécédent mental et un conséquent physique, car rien n'exclut que cette correspondance ne soit qu'une coïncidence. Même si le modèle de la satisfaction permet

1. Russell B., « On the Notion of Cause », (1912), dans *Mysticism and Logic*, Londres, Unwin Paperbacks, 1986, p. 173.

de construire des schémas de prédiction corrects, ce n'est pas un bon modèle pour décrire l'action humaine, ou justifier rationnellement l'opposition évidente entre ce que nous faisons et ce qui nous arrive, entre l'action et l'événement subi.

Pour sauver le modèle de la satisfaction, affaibli par cet argument, certains philosophes ont proposé des variantes moins rudimentaires de ce modèle. Au couple antécédent mental-conséquent physique, ils ont substitué l'idée d'une sorte d'opération continue, à partir d'un programme fixé au préalable ou, plus subtilement, de contrôle continu par ajustements et rectifications constantes à partir d'un programme ouvert et relativement indéterminé. C'est sur ces tentatives de sauver le modèle de la satisfaction que porteront plus précisément mes objections. J'essaierai de montrer qu'elles sont loin d'éliminer la difficulté principale du modèle de la satisfaction qui est, me semble-t-il, la suivante : rien de ce qui précède l'action ne peut servir à l'identifier comme action.

C'est un argument dont on trouve les traces dans la tradition contextualiste, interprétationniste ou anti-causaliste. Il ouvre l'espace d'une théorie rivale que j'appelle *modèle de la réflexion*. J'essaierai de justifier l'idée que le choix de ce modèle n'est pas sans incidence sur la conception que nous devons nous faire des relations entre philosophie de l'esprit et philosophie de l'action. Ce sera ma conclusion. Je me proposais initialement de montrer que cette opposition entre satisfaction et réflexion n'était peut-être qu'un avatar de l'antique dichotomie entre *poiesis* et *praxis*. Mais au point où j'en suis, je préfère ne pas m'embarquer dans cette histoire. J'espère que les aristotélisants ne seront pas trop déçus et que les allergiques au Stagirite seront satisfaits, si je puis dire.

1

Dans le modèle de la satisfaction, les aspects dits mentaux de l'action sont traités comme des variables indépendantes et les comportements comme des variables dépendantes. Je veux dire par là que, dans le modèle de la satisfaction, on s'intéresse aux modifications de la valeur de la variable « comportement » pour toute modification de valeur de la variable « état mental » et non à l'inverse. En fait, le dispositif qui doit rendre opérationnel le

modèle de la satisfaction est assez difficile à construire. En premier lieu, il faut s'assurer qu'on a les moyens de différencier les états mentaux, entre eux, c'est-à-dire de donner leurs conditions d'identité ou des critères de discernabilité. Faute de quoi, il sera impossible d'assigner des valeurs à la variable « état mental », de quantifier sur les états mentaux. En second lieu, il faut en faire autant pour la variable « comportement ». Il n'est pas nécessaire d'être pathologiquement animé par l'esprit de contradiction pour envisager toutes les difficultés qui surgiront à cette étape aussi. Qu'est-ce qui nous permet de dire que deux comportements sont identiques ou différents ? L'indiscernabilité photographique comme le dit Ryle suffit-elle ? Mais les exemples sont innombrables de comportements physiquement indiscernables qu'il convient néanmoins de traiter comme des comportements différents, en raison de la place qu'ils occupent dans une certaine histoire ou un certain contexte. Par ailleurs, rien ne semble exclure que deux comportements physiquement indiscernables soient des expressions d'intentions ou de sentiments différents. S'agit-il néanmoins des mêmes comportements ? Il n'est pas facile de trouver de bonnes réponses à ces questions.

En troisième lieu, il faut arriver à démontrer qu'il existe des relations significatives entre les valeurs des deux variables et pouvoir leur assigner une direction : en l'occurrence, de l'état mental au comportement. Là encore les difficultés sont innombrables.

Vous remarquerez que les conditions requises pour justifier le modèle de la satisfaction ne comprennent pas l'exigence de réduire les états mentaux à des états physiques. Le problème du réductionnisme physicaliste — dit de la causalité oblique — est esquivé dans le modèle de la satisfaction. C'est un modèle épistémologique, c'est-à-dire un modèle qui propose la meilleure explication possible à un certain niveau d'analyse plutôt qu'un modèle ontologique qui indique en quels termes ultimes l'explication devrait être donnée. Cependant, même si nous acceptons de réduire nos prétentions ontologiques, même si nous réussissons à trouver des solutions acceptables aux problèmes de l'identification des valeurs des variables et de l'orientation de la relation, nous n'aurons pas éliminé, pour autant, le redoutable problème des chaînes causales déviantes, dont j'ai donné tout à l'heure une

définition générale et que je vais examiner à présent. On a pris l'habitude de l'illustrer au moyen d'exemples macabres, le plus connu étant celui que Davidson rapporte dans *Actions et événements*¹.

Un homme veut essayer de tuer quelqu'un en lui tirant dessus. Il manque sa victime d'un bon kilomètre mais le coup provoque la débandade d'une horde de sangliers qui piétinent la victime et la tuent². L'homme avait l'intention de tuer quelqu'un. L'état de choses qu'il avait l'intention de réaliser s'est réalisé du fait de son initiative car s'il n'avait pas tiré, sa victime ne serait pas morte. Disons-nous pour autant qu'il l'a tuée intentionnellement ? Établir une relation de correspondance entre l'intention préalable et la réalisation de l'intention est-il suffisant ? Davidson pense que non et il a bien raison, comme d'habitude. En fait, il existe des objections externes et des objections internes à l'idée de satisfaction dans sa version élémentaire, et Davidson insiste plutôt sur les objections internes.

On pourrait dire que les objections externes portent sur tout ce qui peut se passer en dehors de l'esprit de l'agent. Supposons par exemple que l'assassin potentiel ait été un peu moins maladroit, et que, sans rater sa victime, il l'ait blessée seulement. Néanmoins la victime est morte sur le champ en raison d'une faiblesse constitutionnelle, alors que n'importe quelle victime normalement constituée aurait survécu dans des conditions identiques. En d'autres termes, dans des conditions normales, toutes autres choses étant égales par ailleurs, la blessure n'était pas mortelle. On peut prétendre, cependant, que la victime est morte du fait du coup de fusil. Et comme il ne s'est pas produit, entre le coup de fusil et la mort, un événement aussi extraordinaire que le déferlement d'une horde de sangliers hystériques rendant la chaîne causale manifestement déviante, on risque d'être très embarrassé. Disons-nous que l'action a été exécutée de la manière souhaitée ? Que la chaîne causale était appropriée ? Que l'action était intentionnelle ? Nous aurons du mal, probablement, à nous faire une opinion. Nous penserons peut-être que l'assassin n'a simplement pas eu de chance. S'il avait tiré sur une victime un

1. Davidson D., *op. cit.*, 1993.

2. *Id.*, p. 114.

peu plus robuste, ce n'est pas un homicide qu'il aurait sur le dos, mais une tentative d'homicide, ce qui n'est tout de même pas la même chose.

À présent, supposons que la victime soit morte de frayeur en voyant l'assassin tirer dans sa direction, c'est-à-dire qu'elle soit morte juste avant l'impact de la balle. Pourra-t-on défendre l'assassin en disant que cette mort est un effet non voulu de son intention de tuer sa victime et que, par conséquent, il n'est pas responsable de cette mort telle qu'elle a eu lieu ? Bref, pourra-t-on dire que l'assassin n'a pas provoqué intentionnellement la mort de sa victime, parce qu'il n'avait pas l'intention de la tuer exactement de cette façon ?

En réalité, il est difficile de déterminer précisément ce qui fait qu'une déviation est trop importante pour que l'action ne puisse plus être mise au compte de l'agent. Mais, d'un autre côté, il est absurde d'envisager la possibilité que l'agent ait fixé au préalable toutes les modalités d'exécution de son acte. Comment peut-on, dans ces conditions, affirmer que l'action s'est déroulée ou ne s'est pas déroulée de la manière souhaitée ? Si, comme on peut le penser raisonnablement, toutes nos intentions préalables sont relativement indéterminées, quel sens y a-t-il à vouloir fixer précisément l'écart à ces intentions ?

Quoi qu'il en soit, nous ne pouvons pas exclure la possibilité qu'à un moment donné, fut-il extrêmement bref, l'action ait pu cesser d'être intentionnelle. Dans le cas d'actions complexes, d'actions impliquant un patient, d'actions à retardement, il est même absurde d'envisager la possibilité que l'action ait été entièrement du fait de l'agent. C'est pourquoi d'ailleurs, il entre toujours un élément fictif ou attributif dans l'imputation de responsabilité comme Thomas Nagel l'a très bien montré¹.

Les objections dites internes ont de quoi nous rendre encore plus perplexes. Davidson pense qu'elles sont insurmontables. Elles portent sur ce qui est censé se passer dans la tête de l'agent. On les illustre généralement des exemples du type suivant : au moment où l'assassin vise sa victime, la pensée qu'il va la tuer le rend très nerveux et c'est cette nervosité, plutôt que son inten-

1. Nagel T., « La fortune morale », *Questions mortelles* (1979), trad. P. Engel et C. Engel, Paris, P.U.F., 1983, p. 38-53.

tion de la tuer, qui cause, en fait, son acte d'appuyer sur la gâchette. Peut-on dire qu'il a tué sa victime de la manière souhaitée ? La chaîne causale est-elle appropriée ? Est-il raisonnable de dire qu'il a agi intentionnellement ?¹

Ce qu'on peut conclure de ces objections, c'est qu'il est absurde de réduire l'action à une relation de correspondance entre ce qui était souhaité et ce qui a eu lieu. Autrement dit : on ne peut pas se contenter du modèle de la satisfaction dans sa version élémentaire, relation entre deux événements seulement, un antécédent mental et un conséquent physique, parce que rien n'exclut que la réalisation de l'état de choses souhaité soit le dernier maillon d'une chaîne causale déviante.

À la différence de Davidson, Searle estime que le problème des chaînes causales déviantes n'est pas insurmontable². Il est possible, d'après lui, de sauver le modèle de la satisfaction, en introduisant entre l'intention et l'action une variable supplémentaire. En fait, Searle veut décrire l'action en isolant trois variables au lieu de deux, trois variables dont les relations seraient transitives. Les séquences s'enchaîneraient de la façon suivante : nous formons une intention préalable ; cette intention cause une intention en action, laquelle cause l'action elle-même. Lorsque l'intention préalable a causé l'action sans la médiation de l'intention en action, ce n'est pas à proprement parler une action. Car une action qui n'a pas été causée par une intention en action est du genre de celles qui ont suivi une chaîne causale déviante. Mais qu'est-ce qu'une intention en action ? En quoi se distingue-t-elle de l'intention préalable ou de l'action elle-même ? L'intention en action n'est-elle pas une variante de la volonté sous une appellation moins compromettante ? On peut se faire une assez bonne idée de ce que Searle veut dire à partir des exemples qu'il utilise. « Supposez que j'aie l'intention préalable d'aller en voiture à mon bureau. Pendant que j'exécute cette intention préalable, il se peut que j'accomplisse une série d'actions accessoires pour lesquelles il n'est pas nécessaire que j'aie formé d'intentions préalables : ouvrir la porte, démarrer, appuyer sur la pédale d'embrayage, etc. Au

1. Davidson D., *op. cit.*, 1993, p. 114-117.

2. Searle J., « L'intentionnalité de l'action et de l'intention », trad. J. Proust, *Critique*, octobre 1980, 401, p. 990-1010.

moment où j'ai formé l'intention d'aller en voiture au bureau, il se peut que je n'aie pas pensé à ces actions accessoires. Pourtant de telles actions sont intentionnelles. Dans ce cas, j'ai une intention en action, mais pas d'intention préalable¹. »

L'intention préalable, telle que Searle la définit, ressemble à une sorte de déclaration d'intention qui détermine un objectif général et donne de cette façon ses conditions de satisfaction d'ensemble. D'après lui, l'intention préalable est sui-référentielle en ce sens qu'elle ne fixe pas seulement ce qui compte comme réalisation de l'intention, mais demande que ce qui est réalisé le soit en raison de cette intention même. Searle compare la relation de l'intention préalable à l'action à la relation entre un ordre et son exécution. Si quelqu'un m'ordonne de quitter cette salle et que je réponds « j'allais justement partir », avant de quitter la salle, j'aurais fait ce qu'on m'a demandé mais je ne l'aurais pas fait parce qu'on me l'a demandé. En réalité, je n'aurais pas obéi à l'ordre. L'ordre est sui-référentiel en ce sens que ses conditions de satisfaction n'exigent pas seulement la réalisation de l'action ordonnée, mais aussi qu'elle soit effectuée pour obéir à l'ordre².

Passons à l'intention en action. Searle pense qu'on peut caractériser l'intention en action phénoménalement. L'intention en action est une expérience d'action. On peut la mettre en contraste avec ce sentiment d'étrangeté qu'on éprouve, ou pourrait éprouver, si on nous faisait faire quelque chose en appliquant une électrode sur l'aire motrice de notre cerveau. D'après Searle, la réponse des patients qui subissent ce genre d'expérience est invariablement : « Ce n'est pas moi qui l'ai fait, c'est vous ». Pourquoi ? Parce qu'il y a, dit-il, une différence phénoménale évidente entre le cas où on bouge la main et le cas où on la voit bouger indépendamment de ses intentions, les deux étant ressentis différemment par le sujet ou le patient³. Pour ceux qui ne seraient pas convaincus par cet argument phénoménal, Searle ajoute une clause logique. Il pense que des concepts comme « essayer », « réussir », « échouer » s'appliquent aux expériences d'action, sans qu'ils aient d'application analogue dans le cas où cette expé-

1. *Id.*, p. 995.

2. *Id.*, p. 996.

3. *Id.*, p. 1000.

rience est absente, c'est-à-dire, dans l'exemple précédent, lorsque le patient voit sa main bouger, sans qu'il ait eu l'intention de la faire se mouvoir. On ne peut pas dire que le statut de ces expériences d'action soit très clair. Searle s'empresse de préciser, pour le cas où nous serions tentés d'y voir un avatar de la volonté, que ces expériences d'action ou ces intentions en action n'ont absolument rien à voir avec ce que « certains philosophes ont appelé des volitions ou des actes de volonté¹ ». Ce qui montre, en passant, que le fantôme de Ryle n'a pas fini de hanter la philosophie analytique de l'action.

Quoi qu'il en soit, Searle estime que le problème des chaînes causales déviantes n'apparaît que lorsque l'intention préalable n'a pas eu pour effet de causer une intention en action². C'est l'existence d'une intention en action qui garantit non seulement que l'intention a été satisfaite mais qu'elle l'a été par une chaîne causale appropriée. Autrement dit, l'intention en action garantit non seulement que la fin a été voulue, mais aussi que tout ce qui s'est passé pour qu'elle soit réalisée l'ait été également. Mais Searle n'apprécierait probablement pas ce genre de paraphrase où le terme volonté est utilisé généreusement.

C'est vers Alvin Goldman qu'il faudrait se tourner pour avoir une sorte de version « volitionnelle » du schéma de Searle³. Goldman oppose une volonté initiale, qui donne l'impulsion et couvre pour ainsi dire l'ensemble de l'action et des actes de volonté particuliers qui couvrent chaque partie de l'action. Il estime que la volonté en action n'est rien d'autre qu'une suite de séquences volitionnelles qui, chacune, précède une des séquences de mouvement corporel. Si un mouvement est une action, on doit pouvoir le décomposer en une suite de séquences volition-exécution. Toutes les parties d'une action sont donc voulues au même titre que l'action entière. La seule différence entre les volitions initiales et les volitions en action, c'est que ces dernières sont inconscientes ou plus exactement qu'elles sont généralement imperceptibles dans les actions complexes, rapidement exécutées.

1. *Id.*, p. 999.

2. *Id.*, p. 1003.

3. Goldman A., « The Volitional Theory Revisited » dans Brand M., et Walton D., eds, *Action Theory*, Reidel, Dordrecht, 1976, p. 67-84.

Ryle pensait avoir discrédité définitivement l'idée de volonté en affirmant qu'elle nous obligeait à penser que, pour réciter une fable de La Fontaine par exemple, il fallait aussi vouloir en réciter chaque partie, faute de quoi les composants d'une action volontaire seraient involontaires¹. Pour les amis de l'idée de volonté, tels que Ryle les décrit, l'acte volontaire devrait être en principe une conjonction de petits actes volontaires, position qu'il n'est pas difficile de ridiculiser en demandant par exemple : combien d'actes de volonté pose-t-on en récitant le Corbeau et le Renard ? En conduisant une voiture ? En nageant ? En plongeant ? On pourrait répondre à Ryle qu'il n'est pas exclu, en principe, qu'une conjonction d'actes involontaires soit elle-même volontaire, de la même façon qu'une conjonction d'actes stupides ou irrationnels peut-être rationnelle.

Goldman de son côté estime qu'il n'est pas du tout ridicule de postuler l'existence d'une infinité de volitions en action. Il suffit de supposer qu'elles ont été conscientes dans l'apprentissage et qu'elles sont devenues inconscientes avec l'exercice, ce qui ne les empêche pas d'exister. Et Goldman rejette les sarcasmes de Ryle, en suggérant qu'on aurait autant de mal à décompter les désirs que les volitions, ce qui ne nous empêche pas d'affirmer l'existence des désirs. D'après Goldman, bien qu'une volition initiale domine la sphère de l'attention et donne la signification intentionnelle à l'action dans son ensemble, un contrôle permanent, impliquant information et volition, gouverne le mouvement pendant qu'il se déploie, à un degré d'attention moindre. Il s'agit de processus qui ne sont pas « focalisés ». Les pensées ou les volitions qui orientent les mouvements sont marginales ou périphériques mais elles ne sont pas absentes². Ryle opposait le savoir-que et le savoir-comment, en donnant l'exemple de la conduite d'une bicyclette. Nous savons comment faire, mais nous ne savons pas expliquer comment nous le faisons³. Goldman semble estimer que cette différence n'est pas essentielle. Au terme de l'apprentissage, le savoir-que est devenu imperceptible : nous ne savons plus l'exprimer verbalement mais il ne laisse pas d'exister.

1. Ryle G., *op. cit.*, 1978, p. 69-70.

2. Goldman A., *op. cit.*, 1976, p. 74.

3. Ryle G., *op. cit.*, 1978, ch. II.

Je voudrais montrer à présent que ni la théorie de Searle ni la théorie de Goldman ne permettent de sauver le modèle de la satisfaction c'est-à-dire de proposer une solution satisfaisante au problème des chaînes causales déviantes.

La théorie de Searle présente deux inconvénients. En introduisant entre l'intention préalable et le mouvement corporel une variable intermédiaire, Searle ne fait que repousser la difficulté. Car on peut imaginer des chaînes causales déviantes allant de l'intention en action au mouvement corporel.

D'autre part, il affaiblit son argument en donnant pour critère d'identification de l'intention en action une certaine impression ressentie par l'agent. Je ne vois pas du tout comment les intentions en action ainsi définies peuvent nous aider à résoudre le problème des chaînes causales déviantes. Avons-nous des raisons de penser que la différence entre tuer sa victime d'un coup de fusil et la faire mourir parce que le coup de fusil a provoqué l'hystérie d'une horde de sangliers est une différence d'impression ressentie par l'agent ? Il est encore plus difficile d'envisager une solution à la Searle au problème de Davidson, c'est-à-dire à l'objection interne. Avons-nous des raisons de supposer qu'il peut exister une différence phénoménale dans le cas où une pensée inconsciente trouble l'agent de telle sorte qu'il fait involontairement ce qu'il voulait faire volontairement ? Non, puisque c'est précisément la conscience de cette différence qui lui manque. L'agent croit que tout s'est passé de la manière souhaitée, mais il se trompe ; il ne sait pas, il ne s'est pas rendu compte, il n'a pas eu l'impression qu'une détermination involontaire s'est insinuée dans la chaîne qui va de son intention à son action. Bref, on pourrait dire de l'action ce que des théoriciens dits externalistes tels que Putnam ont dit de la signification ; elle n'est pas dans la tête de l'agent, ni dans ce qu'il croit, ni dans ce qu'il ressent.

La théorie de Goldman, elle, ne présente qu'une difficulté, identique à la première difficulté qu'on peut déceler dans l'analyse de Searle. En ajoutant aux volitions préalables une série de volitions imperceptibles, on ne fait que repousser la difficulté, car on peut envisager la possibilité de chaînes causales déviantes allant de chacune de ces infimes volitions à chacune des fractions

de mouvement corporel qu'elles orientent. Rien n'exclut la possibilité qu'il n'y ait en fait qu'une suite très heureuse de coïncidences. Ce n'est pas parce que le mouvement volontaire entier a été décomposé en une infinité de petits mouvements volontaires que la difficulté causale s'est décomposée elle aussi. En réalité, il semble que le problème des chaînes causales déviantes ne soit qu'un cas particulier d'une difficulté plus générale. C'est celle qu'Harry Frankfurt a fort bien diagnostiquée. Cette difficulté est la suivante : le fait que certaines causes sont à l'origine de l'action n'a rien à voir avec les éléments qui la définissent comme action¹. Harry Frankfurt ne s'est pas contenté de cette thèse négative. Il a proposé une thèse positive également, un moyen de classer les mouvements corporels en actions et événements, sans faire intervenir un élément extérieur, une différence à un moment antérieur entre un autre couple d'événements.

Je suis d'accord avec le diagnostic de Frankfurt, mais pas avec la solution qu'il propose, en grande partie parce qu'elle nous ramène aux difficultés des théories de Goldman et de Searle, et en soulève une autre encore, celle qu'on pourrait appeler : problème de l'instrumentalisation du corps.

Frankfurt reconnaît, avec la plupart des philosophes analytiques, que le problème de l'action consiste à expliquer la différence entre ce que fait un agent et ce qu'il subit, entre les mouvements corporels qu'il dirige et ceux qu'il ne contrôle pas. Et il rejette toutes les solutions qui consistent à classer certains mouvements corporels comme actions, uniquement sur la base de leur origine, de leurs antécédents ou d'un événement passé, quel qu'en soit le genre (mental ou physique)². Harry Frankfurt admet que les actions sont susceptibles d'avoir des causes comme n'importe quel événement, mais il pense que le fait d'avoir des antécédents causaux de quelque type que ce soit ne fait pas partie de la nature de l'action³. D'après lui, le fait qu'un événement soit une action n'implique nullement qu'il ait été causé et encore moins qu'il ait des antécédents causaux d'un type spécifique,

1. Frankfurt H., « Le problème de l'action » (1978), dans Neuberger M., éd., *op. cit.*, 1991, p. 241-251.

2. *Id.*, p. 243.

3. *Id.*, p. 242.

mentaux par exemple. Le principal reproche qu'il adresse aux théories causales de l'action, c'est qu'en concentrant leur attention sur ce qui s'est passé *avant* l'action, sur son origine ou ses antécédents, elles semblent suggérer « que la distinction entre actions et simples événements ne saurait se faire au moyen de ce qui existe et se passe au moment où ces événements se produisent [et] qu'il faut se référer à quelque chose de tout à fait extérieur, à savoir la différence à un moment antérieur entre un autre couple d'événements¹ ». Les théories causales laissent supposer qu'une personne accomplit une action non pas parce qu'elle a conscience de ce qui est en train de se passer, mais par ce qu'elle est au courant des conditions antécédentes causant son comportement actuel, ce qui est absurde.

Il me semble qu'on peut admettre, avec Frankfurt, qu'il est absurde d'essayer de classer les mouvements corporels en actions et événements sur la base de leurs antécédents. Mais quelle est la solution de rechange ? Nous devons trouver les moyens de distinguer ces mouvements corporels au moment même où ils se produisent, sans tenir compte de leur origine. Mais il est évident que les mouvements corporels ne révèlent pas d'eux-mêmes s'il y a eu action ou pas, car un même mouvement corporel peut correspondre à une action ou à un simple événement. Autrement dit, on ne peut pas se contenter non plus du critère ostensif (ce qu'on voit, ce que l'action montre). Frankfurt rejette très raisonnablement la solution simple qui consiste à dire qu'on *voit* la différence, même si on a du mal à la décrire, qu'on verrait, par exemple, que certaines actions très compliquées telles que jouer du piano ne sont pas des événements subis. En effet, la complexité des mouvements n'indique pas nécessairement qu'il y a action, le cas des mouvements complexes mais compulsifs des épileptiques étant un contre-exemple évident. Ce qui compte d'après Frankfurt, c'est moins la complexité des mouvements que le fait que l'agent contrôle et dirige les mouvements de son corps². Dans l'exemple du pianiste, si nous jugeons qu'il y a action, ce n'est pas en raison de la complexité du mouvement, mais parce que nous avons du mal à imaginer que ces mouve-

1. *Id.*, p. 242.

2. *Id.*, p. 245.

ments complexes auraient pu se produire si le pianiste n'avait ni dirigé, ni contrôlé ses mains et ses doigts. Bref, la complexité des mouvements corporels est un indice de l'action quand elle nous amène à supposer que l'agent contrôle et dirige les mouvements de son corps. Et Frankfurt conclut en proposant le critère d'identification de l'action suivant : « lorsqu'on est en présence d'un mouvement dirigé et contrôlé par la personne, il s'agit d'une action, quels que soient les antécédents causaux expliquant ce mouvement. Il est possible que la personne agisse bien que son mouvement corporel soit dû au hasard. Inversement, lorsqu'elle ne dirige pas et ne contrôle pas ses mouvements, elle n'agit pas, même si elle est à l'origine des causes antécédentes — les désirs, croyances, intentions, décisions, volitions et tout ce que vous voudrez — dont ces mouvements sont les effets¹. »

Cette thèse rappelle partiellement celle de Searle et de Goldman. Et on peut se demander si Frankfurt n'est pas victime de la même illusion. N'a-t-il pas transféré le problème causal de la relation entre l'intention préalable et l'action à la relation entre les intentions partielles gouvernant le mouvement corporel et le mouvement corporel lui-même ? N'y a-t-il pas, là encore, identification de l'action à partir de ses antécédents ? Frankfurt est bien conscient de la difficulté et c'est pourquoi il insiste sur ce qui sépare sa théorie des théories causales plus ou moins amendées, du genre de celle de Goldman. Frankfurt distingue le mouvement corporel finalisé, dirigé et contrôlé par la personne, des régulations non dirigées par la personne, telles que le mouvement de dilatation des pupilles lors d'une diminution de la luminosité. Et il estime que si on est bien obligé d'expliquer le mouvement corporel finalisé, dirigé et contrôlé en termes de mécanismes causaux, cela ne signifie pas qu'on soit aussi dans l'obligation d'adopter une théorie causale de l'action². D'après Frankfurt, le fonctionnement pertinent de ces mécanismes n'est pas antérieur aux mouvements qu'ils dirigent ; il est concomitant. Je ne crois pas que cet argument soit tout à fait convaincant ou qu'il suffise à éliminer la difficulté rencontrée chez Goldman. En réalité, la principale différence entre la théorie de Goldman ou de Searle et

1. *Id.*, p. 245.

2. *Id.*, p. 247.

celle de Frankfurt, c'est qu'au lieu de faire faire le travail à des volitions fractionnées, si on peut dire, qui prendraient le relais de la volonté initiale, ou de l'intention préalable, il rapporte l'activité à la personne, en évitant cependant de parler de « causalité de l'agent », par opposition à la causalité de l'événement (mental ou physique). En quoi il se montre très sage, car l'idée de causalité de l'agent ne poserait pas moins de problèmes, de son point de vue, que celle de causalité de l'événement. En effet, l'agent peut être placé au début de la chaîne causale allant jusqu'à l'action. Mais la personne, elle, ne peut pas l'être, pour la bonne raison que la personne ne peut être caractérisée selon Frankfurt qu'en termes réflexifs¹.

D'après Frankfurt, tous les agents ont des structures de préférences. Il y a des choses qu'ils désirent et d'autres qu'ils ne désirent pas, des choses qui les attirent et d'autres qui leur répugnent. Selon Frankfurt, il y a aussi des raisons de penser que tous les agents ont des croyances relatives aux moyens de satisfaire ces désirs, et rien ne nous autorise à ne pas leur attribuer la capacité de choisir ou même de délibérer. Cependant, toujours d'après Frankfurt, seul l'agent humain a la capacité d'évaluer ces désirs. Il y a plusieurs façons de les évaluer. Frankfurt se contente de dire que l'agent humain a la capacité de former des désirs et des croyances de second ordre, c'est-à-dire non seulement d'avoir des désirs et des croyances relatives aux moyens de les satisfaire, mais aussi la capacité de juger que certains désirs sont désirables et que d'autres ne le sont pas ou que certaines croyances sont acceptables et que d'autres ne le sont pas. Ce qu'il entend par désir « désirables » n'est pas très clair. Tantôt il semble dire que l'agent humain a la capacité de se forger de nouveaux désirs plus acceptables, par la critique de certains désirs immédiats². Tantôt, il se contente d'affirmer que l'agent proprement humain est celui qui se demande s'il accepte ses désirs, s'il s'identifie avec eux, s'il les endosse, ou si, au contraire, il regrette de les avoir et les rejette³. Il me semble que la seconde variante

1. Frankfurt H.G., « La liberté de la volonté et la notion de personne » (1971), dans Neuberger M., éd., *op. cit.*, 1991, p. 253-269.

2. *Id.*, p. 257.

3. *Id.*, p. 263.

(c'est-à-dire le désir de second ordre en tant qu'identification ou absence d'identification au désir immédiat) est plus raisonnable, car elle ne risque pas d'être ruinée par le problème de Pascal, celui des possibilités que nous aurions de nous forcer à croire, à aimer, à désirer quelque chose que nous ne croyons pas, n'aimons pas, ne désirons pas naturellement. Il semble aussi absurde d'exiger de quelqu'un qu'il nous aime ou qu'il nous désire, que d'exiger de nous-même que nous aimions, désirions ou croyons quelque chose. Et c'est pourquoi il me semble qu'il vaut mieux s'en tenir à la seconde définition de Frankfurt, qui est évaluative, réflexive, rétrospective et non causale. C'est d'ailleurs à cette définition qu'il s'en tient lorsqu'il aborde la question de la personne. Frankfurt estime que, dans la formation de ces désirs et ces croyances de second ordre, se manifeste une capacité d'auto-évaluation réflexive qui est la marque même de l'agir humain. Il pense aussi que c'est en exerçant cette capacité que se forme la personne ou que la personne n'est rien d'autre que ce qui apparaît dans l'exercice de cette capacité.

3

Charles Taylor reconnaît l'importance des thèses de Frankfurt¹. Mais il lui reproche de ne pas être allé assez loin. Frankfurt se serait contenté de proposer une théorie utilitariste de l'activité d'évaluation des désirs et des croyances. Or Taylor estime que rien nous interdit d'attribuer des capacités d'évaluation utilitaire des désirs et des croyances à tous les êtres animés. Ce qui distingue l'être humain, ce n'est pas sa capacité d'ordonner instrumentalement ses désirs et ses croyances, mais sa capacité d'évaluer ses désirs et ses croyances relativement au genre de vie qu'il veut mener, au genre de personne qu'il veut être. L'être humain n'évalue pas seulement ses désirs et ses croyances en termes instrumentaux. Il les évalue aussi en termes de hauteur ou de bassesse, de noblesse ou d'ignominie. L'être humain est celui qui, par exemple, s'efforce de modérer ou réfréner tel ou

1. Taylor C., « What is Human Agency ? » dans *Human Agency and Language. Philosophical Papers*, Vol. I, Cambridge, Cambridge University Press, 1985, p. 15-44.

tel désir parce que ce désir lui paraît minable, mesquin, commandé par la jalousie et l'envie. Il le contrôle parce que la vie qu'il veut mener, la personne qu'il veut être, ne devrait pas, d'après lui, agir en fonction de ce genre de motifs.

Tous les agents, tous les membres d'espèces animées calculent, délibèrent consciemment ou inconsciemment des moyens de rendre leurs désirs compatibles ou de maximiser leur bien-être. Bref, les agents en général semblent tout faire pour rendre optimal le rapport entre leurs préférences et leurs possibilités. Tous les agents évaluent instrumentalement les fins et les moyens. Cependant, dans l'ensemble des agents, seul l'agent humain a la capacité de s'interroger sur la qualité de ses motivations¹. Seul l'agent humain a la capacité de les évaluer moralement, c'est-à-dire, d'après Taylor, en termes de la vie qu'il convient de vivre et de la personne qu'il convient d'être. Il s'agit, bien sûr, d'un point de vue plus ou moins aristotélicien. Mais rien ne nous interdit de proposer une version plus ou moins kantienne de la thèse de Taylor : tous les agents, tous les êtres animés sont capables d'ordonner leurs préférences, de calculer, et de choisir ce qui leur convient le mieux d'un point de vue étroitement instrumental. Mais seul l'agent humain a la capacité d'évaluer ses motifs en s'interrogeant sur leur degré d'impartialité, d'objectivité, d'universalité.

Mais revenons à Taylor. D'après lui, l'être qui se contente de former des désirs et des croyances de second ordre conformément à ses intérêts égoïstes ou instrumentaux est un évaluateur faible. Mais l'être qui évalue ses désirs et ses croyances en se demandant s'ils sont nobles ou ignobles, bons ou mauvais, relativement au genre de personne qu'il veut être, à la vie qu'il veut mener, celui-là est un évaluateur fort². Seul l'être humain peut être un évaluateur fort, et ce n'est que dans l'évaluation forte que se forme la personne. C'est donc au moyen du critère de l'évaluation forte qu'on doit, d'après Taylor, caractériser l'agir humain en tant qu'il se distingue de l'activité de tous les êtres animés³.

Les amendements de Taylor à la thèse de Frankfurt sont assez judicieux, et ils ont l'avantage de laisser intacte la critique du

1. *Id.*, p. 16.

2. *Id.*, p. 23.

3. *Id.*, p. 43-44.

modèle de la satisfaction que propose Frankfurt. L'important en effet, c'est que Frankfurt nous invite à caractériser l'agir humain non pas en termes de ce qui se passe antérieurement à l'action (désirs, croyances, délibérations) mais en termes de ce qui se passe dans l'action même ou rétrospectivement, c'est-à-dire dans ce mouvement d'évaluation. Il y a tout de même certaines ambiguïtés dans les formulations de Frankfurt, des ambiguïtés qui pourraient nous laisser penser que s'il s'est, en effet, débarrassé du modèle causal de l'action au sens strict, il a néanmoins importé l'une de ses difficultés principales en faisant du corps une sorte d'instrument pour la personne.

Nous avons vu, en effet, que Frankfurt distinguait les mouvements d'auto-régulation tels que le mouvement de dilatation des pupilles lors d'une diminution de la luminosité, mouvement dirigé indépendamment de la personne, des mouvements qui sont dirigés et contrôlés par la personne : « Les pupilles se dilatent mais ce n'est pas [la personne] qui fait qu'elles se dilatent, car ce n'est pas elle qui dirige et contrôle ce mouvement. Ce contrôle est dû à un mécanisme qui n'est pas identique à la personne¹. » Mais identifier l'action humaine à partir du degré de contrôle, de direction, d'ajustement des mouvements corporels exercé par une entité non-identique à ces mécanismes, la personne en l'occurrence, c'est s'exposer, semble-t-il, à toutes les critiques qui ont été adressées, dans la tradition philosophique à l'image de l'agent disposant de son corps, pour ainsi dire, ou logé dans son corps « ainsi qu'un pilote en son navire » — selon la formule d'Aristote, reprise par Descartes.

C'est la raison pour laquelle on peut dire, avec Frankfurt, que l'action ne doit pas s'identifier à partir de ce qui s'est passé antérieurement, mais à partir de quelque chose qui apparaît dans l'action elle-même. Mais il ne semble pas que ce quelque chose puisse être le mouvement « finalisé », dirigé ou contrôlé par la personne. Ce qui devrait compter plutôt, c'est la possibilité de pouvoir justifier son action au moment où elle a lieu, en termes de raisons, et de pouvoir l'évaluer en termes de bien ou de mal, ce qui, comme l'a fait remarquer Anscombe, revient à peu près au même².

1. Frankfurt H., *op. cit.*, (1978), p. 245.

2. Anscombe G.E.M., *op. cit.*, 1957.

Néanmoins, Frankfurt, Taylor, Anscombe (et même Davidson, en un certain sens, en dépit de son causalisme apparent) semblent chercher à peu près dans la même direction. Chez tous les auteurs (et chez beaucoup d'autres), le critère d'identification de l'action, ce n'est pas la conformité à des désirs ou des croyances préalablement formés. L'action se définit en tant que redescription ou évaluation simultanée ou rétrospective d'un événement. Elle est identifiée par ce qui apparaît dans l'action ou ce qu'on en dit après coup. En d'autres mots (encore plus rébarbatifs), d'après tous ces auteurs, l'action n'hérite pas ses propriétés de quoi que ce soit d'antérieur : les propriétés de l'action sont conquises dans l'action elle-même. De ce fait, on peut dire que tous adoptent, d'une façon ou d'une autre, un *modèle de la réflexion*.

4

J'ai appelé modèle de la satisfaction celui qui est inspiré de la métaphore du voyage, et je dirai donc que le modèle de la réflexion est celui qui a l'avantage de rendre cette métaphore inopérante, c'est-à-dire d'éliminer un ensemble de problèmes causaux inextricables liés à la description de l'action en tant que succession d'étapes, schéma séquentiel et orienté.

Mais cette métaphore du voyage est puissante. Et nous pourrions avoir des raisons de nous demander si, en l'abandonnant, nous ne nous privons pas de toute possibilité de penser ou de décrire l'action humaine. En réalité nous ne risquons guère de nous retrouver dans cette situation angoissante. Dans nos façons courantes de parler de l'action, nous n'adoptons pas toujours le modèle des étapes, la métaphore du voyage. Nous avons tendance à estimer que le désir peut être raisonné et qu'il peut, pour ainsi dire, accompagner ou suivre la délibération. Il nous arrive aussi de penser que le choix s'exprime dans l'action, de sorte que notre image de l'action est plus globale ou holiste que découpée en étapes. Et même si l'idée peut nous paraître saugrenue à première vue, nous n'excluons pas la possibilité que le verbe « choisir » s'applique à l'action révolue, accomplie, plutôt qu'à ce qui se passe avant l'action, comme lorsqu'on dit, dans une veine stoïcienne, que nous choisissons ce qui nous arrive, ce qui signifie tout simplement que nous acceptons d'en prendre la responsabilité.

D'autre part, c'est très naturellement que nous évitons, parfois, de tenir compte des intentions ou des désirs préalables pour décider du caractère intentionnel ou non d'une action. C'est que nous avons, apparemment, une certaine conscience « ordinaire » de la mortalité des désirs et des intentions, c'est-à-dire que nous savons très bien que, les désirs et les intentions ayant une certaine volatilité, une vocation à disparaître, il se pourrait que des actions accomplies conformément à des désirs ou à des intentions préalables ne soient pas intentionnelles au moment où elles ont lieu.

Je désire très fortement manger des frites, mais ce désir s'évanouit complètement au moment où on m'en sert un plat. Je les mange tout de même, mais ce n'est pas *ce* désir préalable qui fera de mon acte un acte intentionnel. Ce sera, à la rigueur, le désir présent de me conformer à un désir préalable évanoui, ou toutes autres raisons qui auront pour caractéristique de se rapporter à l'action au moment où elle a lieu.

Mais les promesses pourraient peut-être servir de contre-exemples, dans la mesure où il est difficile d'envisager, pour elles, une mortalité similaire à celle des désirs. Supposons qu'ayant envie d'aller voir *Aladdin* au cinéma et ne voyant aucun inconvénient à accompagner mon neveu qui souhaite le voir aussi, je lui promets le lundi d'y aller avec lui le mercredi à 14 heures. Mais le mercredi, vers 14 heures, je n'ai plus envie d'aller voir *Aladdin* et encore moins d'aller le voir avec mon neveu. Cependant, j'y vais tout de même pour tenir ma promesse. Le fait que je tiennne ma promesse sans aucun désir, aucune envie autre que celle de tenir cette promesse, ne vide pas, à première vue, mon acte de son contenu intentionnel. Cependant, une légère modification de cet exemple pourrait suffire à sauver le modèle de la réflexion. Supposons que je promette à mon neveu de l'emmener au cinéma le mercredi à 14 heures pour fêter son anniversaire. Puis j'oublie totalement pour quelles raisons j'ai promis de l'emmener au cinéma ce jour-là tout en me souvenant du fait que j'ai promis de le faire. Je l'emmène donc au cinéma conformément à ma promesse, à cause de ma promesse, pour tenir ma promesse. Mais le fait que je ne sache plus pourquoi je l'ai promis ne prive-t-il pas mon action de son caractère intentionnel ?

Il semble que même dans le cas des promesses, ce qui s'est passé antérieurement à l'acte ne suffit pas à lui donner un caractère intentionnel. Il existe une certaine mortalité des promesses, différente de celle des désirs ou des intentions, mais qui pose des problèmes assez similaires.

Pour conclure, j'aimerais faire une remarque sur ce qu'on peut attendre des relations entre philosophie de l'esprit et philosophie de l'action, si on tient compte de l'opposition satisfaction-réflexion. De nombreux philosophes estiment que le sort de la philosophie de l'action est lié à celui de la philosophie de l'esprit dans la mesure où c'est elle qui, actuellement, est le plus près de nous offrir des théories acceptables des relations corps-esprit ou de justifier rationnellement la psychologie populaire ou la croyance populaire dans l'efficacité causale des désirs et des croyances. Ces espoirs ne seraient justifiés que si toute théorie de l'action devait être construite sur le modèle de la satisfaction. Or j'espère avoir réussi à montrer que cela n'est pas nécessaire.

Par conséquent, même si la philosophie de l'esprit nous apprendrait qu'il est possible d'avoir un accès indépendant à des épisodes mentaux préalables et même si elle était en mesure de nous montrer que ces épisodes mentaux peuvent expliquer causalement leurs conséquents physiques, moyennant une certaine réalisation physique, par exemple, cela ne nous donnerait pas encore les moyens de savoir s'il y a des actions intentionnelles.

En effet, nous pouvons avoir des raisons de penser que ce n'est pas l'existence d'un épisode mental préalable qui peut nous servir à identifier le caractère intentionnel de l'action lorsqu'elle a lieu. Ce qui peut nous aider seulement, c'est savoir si l'agent est en mesure de justifier ce qu'il accomplit. Je ne suis pas du tout sûr que ces arguments suffiront à ébranler, fût-ce partiellement, les convictions des causalistes soutenues et enrichies par la métaphore du voyage. Les malentendus invraisemblables qui ont accompagné la réception du fameux slogan d'Austin « Quand dire, c'est faire » seraient plutôt de nature à me rendre assez pessimiste. Ce sont ces malentendus, très significatifs hélas, que je vais essayer de mettre en évidence, en m'adressant un peu plus directement que je ne l'ai fait jusqu'à présent aux linguistes et sociologues, quitte à revenir sur certains arguments.

3. TROIS ASPECTS DE L'EFFICACITÉ RITUELLE

Dans un article plutôt polémique, Talcott Parsons soutenait qu'il serait assez irrationnel de douter du « pouvoir actif des idées »¹. D'après lui, un tel doute nous obligerait à renoncer à des secteurs entiers d'activité scientifique rentable, si on peut dire, ce qui serait absurde. La théorie économique, telle que Parsons la conçoit par exemple, devrait être abandonnée, dans la mesure où elle repose sur un schéma analytique dans lequel nos idées, au sens large (capacités logiques, savoirs empiriques, jugements moraux et esthétiques), sont censées « influencer » nos conduites. Bien sûr, l'argument de Parsons n'est pas une variation sur le thème de l'action à distance. Parsons ne croit certainement pas au « pouvoir actif des idées » comme d'autres ont la faiblesse de croire qu'il n'est pas impossible de déplacer des armoires normandes par la seule puissance de son esprit ou de sa volonté, considérées comme des forces mentales. Lorsqu'il parle du « pouvoir actif des idées », il pense probablement à quelque chose qui ressemble au modèle de la satisfaction, c'est-à-dire qu'il doit juger qu'il est possible de justifier au moyen d'une loi causale ou, plus modestement, par de bonnes raisons, la succession de certaines expériences.

Quelles que soient les réserves qu'on peut avoir à l'égard des thèses de Parsons en général, il ne semble pas, à première vue, qu'il puisse avoir tort sur ce point particulier. Si une certaine forme de scepticisme ne paraît pas déraisonnable lorsque nous envisageons des actions soustraites, en apparence, aux exigences de la vie sociale (de purs mouvements corporels qui seraient commandés par notre « volonté », par exemple)², ce scepticisme doit être suspendu, semble-t-il, aussitôt que nous nous intéressons à l'échange social.

1. Parsons T., « The Role of Ideas in Social Action », (1938) dans *Essays in Sociological Theory*, Free Press Paperback, New-York, 1954, p. 19-33 (p. 23, en particulier).

2. Lewis H. D., « L'esprit et le corps selon O'Shaughnessy », *op. cit.*, 1986.

En effet, si ce doute nous paraissait justifié, nous serions dans l'obligation de réputer mystérieuses les séquences les plus évidentes de l'interaction : ordres, questions, prières, remerciements, dans lesquels les événements mentaux tels que l'interprétation d'un commandement, ou une hypothèse sur les attitudes d'autrui, semblent être les « causes » d'un mouvement corporel (on nous appelle et nous venons ou nous fuyons, selon les cas ; nous supposons qu'autrui nous veut du mal et nous allons vers lui ou le fuyons selon les cas, etc.).

Nous avons alors, clairement, un antécédent et un conséquent. L'antécédent, c'est l'ordre ou la menace tels qu'ils sont compris (c'est-à-dire des idées, des événements mentaux), et le conséquent c'est le mouvement local du corps (c'est-à-dire un événement physique)¹.

Ainsi, pas plus que nous ne pouvons, semble-t-il, remettre en question un principe aussi central que le principe de non-contradiction dans nos constructions théoriques, nous ne pouvons douter, dans nos constructions empiriques portant sur le comportement humain, de l'existence de relations causales entre événements mentaux et événements physiques, genre dont les espèces bien connues sont : l'efficacité symbolique, l'efficacité rituelle, la définition de la situation, l'interaction stratégique, la désignation, la prophétie auto-créatrice, etc.

À ce niveau de généralité, d'ailleurs, le mélange de variétés hétérogènes n'a peut-être pas tellement d'importance et c'est pourquoi on peut ranger ensemble, des « événements mentaux » plutôt « conatifs » (désirs, préférences) et d'autres plutôt « cognitifs » (croyances, représentations).

Cependant, cette vitalité ou cette centralité de l'idée de « causalité mentale » ne l'empêche pas de demeurer mystérieuse. Il suffit de se demander : « Mais que signifie plus précisément causalité ? » (pour ne rien dire d'autres termes plus embarrassants encore tels qu'« événement mental »), et ces constructions si évi-

1. Wittgenstein nous a recommandé de réfléchir au langage du commandement (le langage impératif) pour comprendre ce que nous appelons activité « volontaire ». Searle s'est largement inspiré de cette recommandation pour justifier sa thèse dite de la causalité « mentale », ou « intentionnelle ». Cf. Searle J., *op. cit.*, 1985, p. 110.

dentes perdent assez rapidement leur justification rationnelle (ce qui ne veut pas dire qu'elles cessent d'être utilisables et indispensables dans la vie courante).

C'est du moins ce que je vais essayer de montrer, pour introduire, en quelque sorte, à mes remarques sur les thèses de J.L. Austin, qui se serait bien gardé, me semble-t-il, de donner une signification causale étroite au slogan séducteur « Quand dire, c'est faire », ou à ses termes évocateurs : performatif, perlocutoire.

On ne peut pas dire que ses nombreux exégètes aient fait preuve de la même prudence (ou de la même sagesse). Les sociologues qui se sont intéressés aux idées d'Austin ont souvent commis l'erreur de leur donner un sens causal inapproprié. En conséquence, lorsqu'ils ont été critiques, leurs objections ont été jugées déplacées et lorsqu'ils ont été enthousiastes, leur soutien a été jugé plutôt encombrant. Il y a, heureusement, quelques exceptions¹.

1

Nous pouvons essayer de donner une définition assez large de la causalité qui devrait en principe pouvoir s'appliquer à toutes sortes d'événements mentaux ou physiques du moment que nous avons des raisons de prétendre qu'ils sont en relation de conjonction constante².

1. Sur ce sujet, voir les objections de François Recanati, dans sa postface à l'édition de l'ouvrage d'Austin, *Quand dire, c'est faire*, Points-Seuil, p. 203, aux arguments de Pierre Bourdieu, exposés dans *Ce que parler veut dire*, Paris, Fayard, 1982, p.105 ; parmi les exceptions, je mentionnerai tout particulièrement l'essai de François-André Isambert, *Rite et efficacité symbolique*, Paris, Le Cerf, 1979.

2. La littérature sur ce thème est immense. On ne peut pas éviter de citer Hume, le provocateur : *Traité de la nature humaine* (1739-1740), trad. A. Leroy, Paris, Aubier, 1983. Après une longue éclipse, le thème de la causalité fait un retour assez réussi, comme le remarque M. Bunge : « The Revival of Causality », dans Floistad G., éd., *Contemporary Philosophy* (2), La Haye, M. Nijhoff, 1982, p. 133-135. Il est difficile de nier l'importance des articles de Davidson dans ce renouveau de la réflexion sur la causalité « mentale », *op. cit.*, 1993. Ricœur donne un excellent compte rendu de la discussion contemporaine (de Weber à Von Wright) qui a dégagé une place pour l'imputation causale « singulière » aux côtés de la causalité « nomologique » (recherche d'une « couverture légale » ou

J'ai examiné, dans les chapitres précédents des façons assez différentes de caractériser la relation causale. En réalité, les critères d'identification de la causalité varient énormément. Il y a, pour ainsi dire, une indétermination de la définition de la causalité (plusieurs définitions différentes se valent).

Dans ce chapitre, je m'en tiendrai à des critères formels : directionnalité, extensionnalité, indépendance logique. J'ai dit des choses très générales des deux derniers. Je profiterai de cette discussion pour approfondir le sujet. Mais ce qui m'importe, c'est de montrer que le terme « causalité » paraît inapplicable aux relations entre les idées et les actions, même lorsqu'on lui donne ce genre de définition minimale. Commençons par le dernier trait, l'indépendance logique, le plus important, peut-être.

La notion d'indépendance logique signifie, en gros, que même si la relation entre deux événements, entre, par exemple, « mettre ma main sur le feu » et « éprouver une sensation de chaleur ou de brûlure », nous semble nécessaire, c'est-à-dire qu'il ne pourrait pas se faire que, dans des conditions normales, je ne sois pas brûlé si je mets ma main sur le feu, cette nécessité n'a rien de logique ou de conceptuel. Car si nous pensons qu'il est possible de parler de nécessité dans ce cas, c'est d'une nécessité *a posteriori* qu'il s'agira, c'est-à-dire d'une nécessité connue au moyen d'une expérience passée et permettant de prédire un événement futur en invoquant l'existence d'une loi.

Le fait que cette prétendue nécessité de la connexion causale est connue *a posteriori*, c'est-à-dire qu'elle se rapporte directement ou indirectement à l'expérience, nous est rendu manifeste au moyen de certaines expériences de pensée. Nous pouvons avoir l'impression d'être en pleine fiction lorsque nous essayons d'imaginer que nous pourrions poser tranquillement une main sur le feu sans qu'elle soit protégée et que la main reste intacte ou encore qu'elle bronze joliment. Mais fiction ne veut pas dire contradiction : des systèmes entièrement fictifs peuvent être cohérents. Et le fait que nous ne nous contredisons pas, à propre-

... d'une « loi » expliquant la succession de certains événements, — tyrannie et révolte, par exemple — introuvable, bien sûr, dans les « sciences humaines »). Cf. Ricoeur P., *Temps et récit*, I, Paris, Le Seuil, 1983.

ment parler, en déclarant qu'une main posée sur le feu ne brûle pas, suffit, semble-t-il, à justifier l'idée que la nécessité de la connexion naturelle n'est pas logique ou conceptuelle car sa négation n'implique pas contradiction. Au mieux, on pourrait dire qu'on se contredit étant donné le reste de nos croyances à propos du feu et des mains ; mais ces croyances nous ramènent à un moment ou l'autre à l'expérience. Les termes d'une relation causale devraient donc être, en principe, logiquement indépendants. La relation causale est non nécessaire en ce sens, même si nous avons de bonnes raisons d'admettre l'idée d'une nécessité d'une autre nature, une nécessité de la nature justement, disant que certains événements physiques s'excluent, et que d'autres s'associent inévitablement.

Passons à l'extensionnalité. Appliquée aux relations causales, cette notion signifie que si un événement en cause un autre, il le cause sous toutes les descriptions de cet événement. Si j'ai détruit la Tour Eiffel, et que la Tour Eiffel est le plus haut monument de Paris, ou le pylône posé sur le Champ de Mars ou le phare de la capitale, alors, en détruisant la Tour Eiffel, j'ai détruit également le plus haut monument de Paris, posé sur le Champ de Mars, ou le phare de la capitale.

Enfin, la notion de directionnalité, appliquée aux relations causales, signifie que cette relation est non symétrique et non réflexive. « Le coup de poignard de Ravailac a causé la mort d'Henri IV », mais « La mort d'Henri IV n'a pas causé le coup de poignard de Ravailac », même si elle a causé d'autres choses. La relation « être cause de » n'est pas symétrique. Par ailleurs, il serait absurde de dire que « Le coup de poignard de Ravailac a causé le coup de poignard de Ravailac » ou que « La mort d'Henri IV a causé la mort d'Henri IV ». La relation « être cause de » n'est pas réflexive, contrairement à la relation « être identique à » par exemple, (je peux être identique à autre chose, tout en étant identique à moi-même).

Pour beaucoup, ces trois façons de caractériser la relation causale, indépendance logique, extensionnalité, directionnalité, sont insuffisantes ou inappropriées. Ainsi, on a soutenu que la notion d'indépendance logique était inadéquate car, pour que deux événements entretiennent une relation causale, il est nécessaire qu'ils appartiennent à une même classe : celle des événements

qui peuvent interagir¹. Du coup, la notion de directionnalité perd de son importance, car si les termes de la relation causale ne sont plus indépendants, on devra parler de changement d'état d'un système plutôt que d'action unilatérale d'un agent sur un patient.

C'est la raison pour laquelle Russell, par exemple, nous recommandait vivement d'essayer d'oublier l'idée de cause, en suggérant, pour nous rassurer, qu'on ne ferait rien d'autre en ce cas que suivre les excellentes habitudes des physiciens contemporains, dont les théories se portent de mieux en mieux depuis qu'elles ont cessé de rendre un culte aux causes, sans renoncer, cependant, au déterminisme².

De toute façon, l'exigence de directionnalité avait déjà été contestée par Aristote, qui insistait sur l'importance de la puissance passive. La flamme ne peut brûler la main que si la main est combustible ou inflammable. La puissance passive de la main, sa combustibilité agit donc à sa façon comme la puissance active du feu. Le dispositif prétendument causal est celui de l'action réciproque et non celui de l'action unilatérale d'un agent sur un patient³.

Enfin, l'idée d'extensionnalité appliquée à la causalité semble elle aussi inappropriée. Car lorsque le choc d'une pierre grise cause la destruction d'une vitrine rouge, la pierre ne cause rien en

1. Cette discussion remonte à Aristote (au moins) qui opposait les théories physiques respectant le principe « le semblable agit sur le semblable » à celles qui préféraient le principe inverse : « le dissemblable agit sur le dissemblable », et concluait en soutenant que l'agent et le patient, dissemblables et contraires par l'espèce, devaient être néanmoins semblables et identiques par le genre. Cf. *De la génération et de la corruption*, I, 7, éd. et trad. C. Mugler, Paris, Les Belles Lettres, 1966, p. 28-32.

2. Russell B., *op. cit.*, 1986, p. 173-199.

3. Les réflexions d'Aristote sur la relation action-passion, agent-patient sont dispersées dans plusieurs écrits : *Physique*, *La Métaphysique*, *L'Éthique à Nicomaque*. *De la génération et de la corruption*. Dans son commentaire, Sarah Waterlow a mis en évidence les principales difficultés et les obscurités de cette analyse. Waterlow S., *Nature, Change and Agency in Aristotle's Physics*, Oxford, Clarendon Press, 1982 ; voir en particulier le chapitre intitulé « Agent and Patient », p. 159-203.

tant qu'elle est grise et la vitrine n'est pas brisée en tant qu'elle est rouge. Ce n'est donc pas sous toutes leurs descriptions que les événements entrent en relation causale, comme on l'a vu¹.

Il y a plusieurs façons différentes de réduire la portée de ces arguments dans le domaine physique. On peut admettre, par exemple, que la prétendue causalité n'est pas directionnelle lorsque nous envisageons des classes d'événements, tout en défendant l'idée que la causalité est orientée lorsque nous considérons des événements singuliers. De plus on peut soutenir que, même si les objets n'ont pas de relation causale sous leur description « colorée », il n'en résulte pas, pour autant, que la relation causale soit intentionnelle. Mais je n'insisterai pas sur ces discussions subtiles ici².

En effet, même si on peut discuter de la valeur des objections anti-causalistes dans le domaine physique, elles semblent imparables, si on peut dire, lorsqu'on les applique au domaine de l'action humaine.

Il est clair que la notion d'extensionnalité n'est pas utilisable dans le domaine de l'action, car autrement nous serions dans l'incapacité de distinguer ce qui dépend de nous et ce qui n'en dépend pas, les conséquences attendues ou inattendues de nos actes, les gestes volontaires et les coïncidences.

Nous n'aurons aucune difficulté à comprendre les tourments d'Œdipe, si nous abandonnons le point de vue extensionnel. Nous admettons qu'Œdipe ne souhaitait pas tuer Laïos, son père, même s'il voulait tuer pour se venger l'homme qui l'avait bousculé à la jonction des routes de Delphes et de Daulis, et, bien que son père ne fût autre que cet homme, plutôt agressif. Œdipe ne tue pas Laïos « sous toutes ses descriptions », bien entendu, mais seulement en tant qu'homme agressif qui veut le pousser contre le talus. Du point de vue causal étroit ou extensionnel, ce genre de coïncidence malheureuse n'aurait aucun sens, car Laïos est mort quelle que soit sa description ! (« Le père d'Œdipe est mort » équivaut à « L'homme agressif rencontré par Œdipe à la jonction

1. Sur ce sujet aussi, Aristote avait anticipé quelques bons arguments contemporains. Cf. Everson S., « L'explication aristotélicienne du hasard », *Revue de philosophie ancienne*, 1, 1988, p. 39-76.

2. Kim J., « Causation, Nomic, Subsumption and the Concept of Event », *The Journal of Philosophy*, 8, LXX, 1976, p. 217-236.

des routes de Delphes et de Daulis est mort. ») Observons, en passant, que certaines théories psychologiques utilisent cet argument extensionnel pour nier que l'intentionnalité soit un trait essentiel de l'action, ce qui n'est pas maladroit — mais tout de même un peu abusif. Si Œdipe s'est laissé entraîner à tuer l'homme qui l'avait bousculé, c'est précisément parce que cet homme était son père, et qu'il est bien utile de se débarrasser de son père lorsqu'on désire ardemment — bien qu'inconsciemment — épouser sa mère. L'intentionnalité de l'action humaine n'est pas un trait essentiel et les coïncidences malheureuses n'existent pas. Ce qui revient au même, dans ce cas.

La notion d'indépendance logique est elle aussi difficile à appliquer, car nous n'avons aucun bon moyen, semble-t-il, d'identifier ou plus exactement d'individuer une croyance ou une intention ou de garantir l'existence indépendante d'une croyance ou d'une intention, sans nous aider des paroles ou des gestes qui en sont les indices ou les expressions. Et ce n'est pas du tout la même chose de dire qu'un acte *exprime* une croyance et qu'un acte est *causé* par une croyance.

De ce fait, la notion de directionnalité est également inapplicable, car de la confusion de l'antécédent et du conséquent dans une relation telle que celle qui unit une intention, une croyance et l'acte qu'elle est censée causer, on peut dériver, sans se couvrir de ridicule, l'idée qu'en réalité, ce sont nos actes qui causent nos croyances ou nos intentions ou nos sentiments, comme W. James l'avait soutenu, ou que c'est la satisfaction qui engendre « rétrospectivement » le désir, comme l'a suggéré Russell¹.

Par ailleurs, il est évident qu'un ordre, une prière, une question ne peuvent être obéis ou satisfaits que s'ils sont, d'une manière ou d'une autre, compris et reconnus par celui qui les reçoit, et si celui qui les reçoit est, d'une manière ou d'une autre, plus ou moins capable ou disposé à leur obéir ou à les satisfaire. Nous trouvons là l'équivalent moral du problème des puissances passives et nous sommes dans l'obligation, semble-t-il, d'admettre que l'idée d'un pur pouvoir des mots est inintelligible.

1. James W., *Précis de Psychologie*, 1908, trad. É. Baudin et G. Bertier, Paris, Rivière, 1929 ; Russell B., *The Analysis of Mind*, (1921), Unwin Paperbacks, 1989, p. 58-76.

Enfin, si on admet l'idée qu'une relation causale lie des événements et non des objets (c'est-à-dire plutôt des mouvements de boules de billard que des boules de billard), les problèmes de localisation spatio-temporelle deviendront pratiquement insolubles dans le cas des événements moraux ou sociaux.

Ainsi, la question « Où et quand Socrate est-il mort ? » ne semble pas dépourvue de sens. De même, la phrase « Xantippe est devenue veuve à cause de ou à la suite de la mort de Socrate », est intelligible. Mais comment pouvons-nous localiser cet événement ? Où Xantippe est-elle devenue veuve ? Dans la prison de Socrate ? À l'endroit où la nouvelle lui fut annoncée ?¹ L'événement moral ou social semble bien être une conséquence, un résultat, un effet d'un autre événement, mais il paraît absurde d'essayer de le localiser spatialement.

D'autres énigmes concernant la localisation d'événements moraux ou sociaux ont une forme légèrement différente. Ainsi, on peut dire qu'une somme d'actions individuelles irrationnelles (des dépenses de luxe par exemple) ont des conséquences, des effets, des résultats rationnels (l'augmentation moyenne du bien-être peut-être). Mais *où* cette conséquence a-t-elle lieu ? S'agit-il d'un événement qui s'ajoute aux événements individuels comme un effet à sa cause ?

Cette difficulté à localiser ou à fixer le lieu où se produit la cause ou l'effet dans les événements moraux ou sociaux sert généralement à justifier l'argument provocateur : certains événements (le bien-être général, par exemple) n'ont pas de cause, ce qui contredit une sorte d'idée *a priori* apparemment inébranlable.

Mais il n'est pas nécessaire d'aller jusque-là : nous pouvons nous contenter d'observer que les termes de causalité ou d'efficacité sont de faux amis lorsqu'on veut les appliquer aux événements moraux ou à l'interaction entre événements mentaux et événements physiques.

Tout ceci pourrait nous inciter à considérer que la prudence dont Austin a fait preuve dans le maniement de la notion de causalité symbolique en général et du pouvoir des mots en particulier, cette prudence devrait être digne d'éloge.

1. Cf. la discussion de Jaegwon Kim, « Noncausal Connections », *Noûs*, 8, 1974, p. 41-52.

Austin a cherché, semble-t-il, à éviter deux écueils, deux abus dans l'interprétation de « Quand dire, c'est faire » ; d'une part, la vacuité de la théorie de l'autovérification ou de la réflexivité du langage et, d'autre part, la fausseté manifeste des descriptions purement causales de l'efficacité rituelle ou du pouvoir des mots.

Si telle est bien sa position, elle mérite d'être étudiée. Le point de vue « ni, ni » comme on peut l'appeler (ni pure réflexivité ni stricte causalité) n'est pas le moins raisonnable dans l'analyse du prétendu pouvoir des mots, même s'il a incontestablement quelque chose de décevant.

2

À première vue, la théorie d'Austin, telle qu'elle fut exposée d'abord dans un texte écrit en 1946, intitulé « Other Minds », puis sérieusement amendée par la suite dans « Performatif-Constatif » (1958) et dans le fameux recueil de conférences *How to do things with words* (1962)¹, cette théorie n'a pas grand-chose à voir avec les problèmes que je viens d'évoquer.

Austin semble surtout soucieux de ruiner les théories de la signification qui réduisent le sens d'un énoncé à ses conditions de vérité. C'est la partie négative de sa théorie, la plus solide peut-être. Mais dans sa partie positive, plus fragile, il donne des indications sur ce qui m'intéresse, sa conception du pouvoir des mots. Avant d'y venir, je vais évoquer rapidement la partie négative de l'argument d'Austin.

Ce qu'Austin attaque en général, c'est la réduction plus ou moins fréguénne du sens d'une phrase à ses conditions de vérité. En gros, cette réduction consiste à évaluer chaque phrase au moyen d'un test. Peut-on se poser à l'égard de cette phrase (de façon sensée) la question est-ce vrai, est-ce faux ? (indépendamment d'ailleurs des moyens dont nous pourrions disposer pour la

1. Austin J.L., « Other Minds » (1946) dans *Philosophical Papers*, Oxford University Press, 2^e éd., 1970, trad. fr. : *Écrits Philosophiques*, op. cit., 1994, p. 45-91 ; « Performatif-Constatif » (1958) dans *La philosophie analytique*, Paris, Minuit, 1962, p. 271-304 et *How to do Things with Words*, Oxford, Oxford University Press, 1962 b, trad. fr. G. Lane, *Quand dire, c'est faire*. Paris, Le Seuil, 1970.

vérifier ou la réfuter). Si le fait de se poser la question n'est pas absurde, la phrase est sensée ; s'il est impossible de se la poser raisonnablement, c'est que la phrase est dénuée de signification. Ainsi on peut se demander à propos de la phrase « La terre est plate », est-ce vrai ? Est-ce faux ? Ou « Est-il vrai ou faux que la terre soit plate ? » Et cette question est sensée, même s'il sera sans doute indispensable de la préciser si on veut en fixer la valeur de vérité (plate comment, comme une feuille de papier à cigarette, par exemple ?). Cependant, on s'aperçoit très vite que ce test est inacceptable, parce qu'il nous contraint de réputer dénués de sens des énoncés qui en ont manifestement pour tout le monde. Personne ne dira que l'impératif « Viens ici ! » ou la question « Quelle heure est-il ? » ne signifient rien. Pourtant, il est évident qu'on ne peut pas évaluer le caractère sensé de « Viens ici ! » ou de « Quelle heure est-il ? » en se demandant : est-il vrai ou faux que « Viens ici ! » ? Est-il vrai ou faux que « Quelle heure est-il ? ».

Ce fait n'a pas manqué d'être souligné par la plupart des grammairiens depuis l'Antiquité. Mais l'originalité d'Austin provient de ce qu'il a cherché à donner la plus large portée à cette observation, en montrant qu'elle pouvait s'appliquer à des assertions telles que : « J'affirme que la terre est plate », « Je te conseille de lui rendre sa perceuse électrique » etc. Lorsque je dis « Je te conseille de lui rendre sa perceuse électrique », on peut me rétorquer « Qui es-tu pour me donner des conseils ? » ou « Ton conseil est inutile, car je l'ai déjà rendue », mais on ne peut pas me répondre « Ton conseil est faux ». Ces énoncés n'ont que l'apparence d'assertions susceptibles d'être vraies ou fausses, ce sont des « imposteurs ». Et Austin sera tenté de dire que toutes nos assertions peuvent être suspectées d'imposture, relativement au vrai et au faux.

Cependant le phénomène qui intéresse Austin, celui de la liberté que semblent prendre certains énoncés à l'égard du vrai et du faux, tout en restant sensés, peut être envisagé de manière tout à fait différente.

On peut soutenir, en effet, que, si la phrase « J'affirme que la terre est plate » n'est pas une assertion ordinaire, c'est-à-dire susceptible d'être vraie ou fausse, c'est tout simplement parce qu'elle est toujours vraie, autovérifiante ou réflexive.

En effet, mon assertion portant sur la forme de la terre peut être vraie ou fausse, selon les cas, fausse, par exemple, lorsque je dis que la terre est plate comme une feuille de papier à cigarette. Mais lorsque je dis : « J'affirme que la terre est plate », ce que je dis n'est pas vrai ou faux selon les circonstances. C'est toujours vrai. Pourquoi ? Mais parce que personne ne peut nier que j'ai fait une affirmation, même si tout le monde en nie le contenu.

On s'aperçoit vite qu'on peut généraliser cette observation. Aussitôt qu'on précise la force ou la valeur qu'on entend donner à une phrase (conseil, avertissement, suggestion, etc.), la phrase devient vraie, nécessairement. Quelle que soit la valeur du conseil, il est vrai que je conseille lorsque je conseille quelque chose à quelqu'un.

En d'autres termes, c'est tout simplement le fait d'expliciter la valeur ou la force illocutoire, comme l'appelle Austin, d'un énoncé, qui lui fait perdre sa valeur descriptive. Car une phrase ainsi préfixée devient toujours vraie, tautologique, et les tautologies ne disent rien, ne décrivent rien, on le sait bien.

Les Jésuites se sont interrogés sur la valeur des compromis moraux qu'offraient les paradoxes de l'autovérification. Ils ont montré, entre autres, que le caractère autovérifiant ou réflexif des phrases préfixées par ce qu'Austin appelle performatif pourrait nous donner des moyens (peut-être condamnables) d'éviter des péchés gravissimes, le mensonge par exemple, en accomplissant des péchés plus bénins : des sortes d'infractions à certaines normes de communication (le choix unilatéral des contextes ou des références, comme le dirait Fauconnier)¹. Ainsi, on peut éviter de mentir en préfixant le témoignage de bonne foi « Je dis la vérité » d'une affirmation qui rend la phrase autovérifiante ou réflexive ou vraie quelles que soient les circonstances. Il me suffit, en effet, de prononcer à voix basse ou en moi-même, silencieusement « Je dis que », puis à voix haute et intelligible « Je dis la vérité », pour éviter le mensonge et le parjure, même si je mens. En effet, en focalisant ma phrase sur le préfixe « Je dis que », j'annule, en quelque sorte, mon engagement à dire le vrai, pour

1. Cf. Fauconnier G., dont l'article m'a beaucoup servi pour ces digressions « jésuitiques ». « Comment contrôler la vérité ; remarques illustrées par des assertions dangereuse et pernicieuses en tout genre », *Actes de la Recherche en Sciences Sociales*, 25, 1979, p. 4-22.

me borner à constater que je le dis, ce qui est manifestement vrai : je n'ai jamais dit que j'allais dire le vrai, j'ai dit seulement que je disais que j'allais dire le vrai, ce qui n'est pas faux.

Cette procédure un peu puérile, dite de la restriction mentale, de l'équivoque ou de l'amphibologie, justifiée au nom du principe de la propriété des pensées, peut être étendue et généralisée.

Un peu d'entraînement nous permettra de concilier notre respect pour certaines autorités morales et notre souci, plus ou moins légitime, de profiter de la vie. Si on me demande : « As-tu péché ? », je peux m'en sortir sans mentir, même si j'ai commis des infractions déplorables, en répondant à voix haute : « Non », tout en ajoutant à voix basse, ou en moi-même, silencieusement : « Pas aujourd'hui, du moins ».

Ces exemples confirment que ces compromis moraux ne sont possibles qu'au prix d'infractions à certaines exigences référentielles ou topiques, généralement non formulées. Mais c'est peut-être aussi à ce prix, également, que des assertions telles que « J'affirme que la terre est plate » peuvent être considérées comme des assertions non descriptives ou non représentatives, mais réflexives ou autovérifiantes. Si on reproche à quelqu'un de s'être trompé en ayant affirmé que « la terre est plate », je ne crois pas qu'on acceptera facilement qu'il se dérobe en disant : « Mais je me suis contenté de dire que j'affirmais quelque chose » (ce qui n'est pas faux).

Toute cette discussion pourrait contribuer à montrer qu'Austin aurait été victime de ses tests, comme ses adversaires vériconditionnalistes. Mais, en réalité, Austin n'a jamais songé à justifier le caractère non descriptif ou non représentatif des énoncés préfixés par un performatif en s'enfermant dans l'argument de la réflexivité ou de l'autovérification. En effet, il dit (et répète) que ces énoncés sont exposés aux échecs, aux malheurs, aux infortunes des rites mal effectués, c'est-à-dire dans des circonstances inappropriées par des personnes non autorisées.

Si Austin s'était contenté de réduire les énoncés dits performatifs à la réflexivité, l'argument aurait été déplacé. Car le propre de l'énoncé autovérifiant ou réflexif, c'est précisément d'être immunisé à l'égard des circonstances et des personnes : il est toujours heureux, malheureusement, pour notre vie morale, comme le diraient ceux qui rejettent formellement les recettes des jésuites.

Austin refuse apparemment de définir les traits non descriptifs des énoncés performatifs en utilisant l'argument de la réflexivité. Les énoncés performatifs ne sont pas non-descriptifs parce qu'ils seraient toujours vrais. Mais a-t-il cherché la source du caractère non représentatif des énoncés performatifs dans une thèse disant qu'ils sont, en quelque sorte, des antécédents du faire, des énoncés ayant vocation à modifier des états de choses plutôt qu'à les constater ? En d'autres termes, a-t-il proposé une version causale de la relation entre dire et faire ? Cela n'est pas très clair, en dépit de ce qu'on dit la plupart du temps (c'est l'interprétation « orthodoxe »). Austin donne souvent l'impression d'opposer ce qu'on fait en disant quelque chose (suggérer, avertir, conseiller, etc.) et ce qu'on fait par le fait de dire (vexer, persuader, flatter, irriter, effrayer, etc.). Mais ce qu'on fait en disant semble contenir un élément causal : nous devons garantir la compréhension du conseil, de l'ordre, de la suggestion.

Austin donne aussi l'impression de penser que ce genre d'initiative peut s'insérer naturellement dans une chaîne causale, dont les premières séquences s'enchaînent nécessairement, et les suivantes, de façon contingente. Aussitôt après avoir déclaré l'ouverture de la séance, un nouvel état de choses s'est créé nécessairement : celui dans lequel la séance est ouverte. Mais, à la suite de l'ouverture, l'assistance peut être ennuyée ou satisfaite : il n'y a là aucune nécessité. La relation entre l'ouverture de la séance et la satisfaction (ou l'ennui) est contingente. Cependant, en tout, on peut tracer une sorte de ligne causale allant de la déclaration d'ouverture aux réactions de l'assistance, car si la séance n'avait pas été ouverte, il n'y aurait pas eu ces réactions et si on n'avait pas déclaré la séance ouverte, elle ne serait pas ouverte. Même si nous rejetons cette version causale fantaisiste, nous pouvons avoir des raisons d'adopter une interprétation causale plus raisonnable, disant qu'Austin conçoit l'énoncé performatif comme un acte visant à transformer le monde, par opposition à l'énoncé constatatif qui se borne à le décrire. Et des exemples tels que « Je déclare la séance ouverte » qui causent, en quelque sorte, directement l'ouverture de la séance semblent renforcer cette interprétation quasiment magique du pouvoir des mots, tel qu'Austin la concevrait.

Cette interprétation causale de la théorie d'Austin semble directement inspirée des thèses d'Anscombe, qui distingue deux relations au monde, selon leur direction d'ajustement¹. Parfois, nos paroles, nos représentations sont censées s'ajuster au monde. Nous disons des choses vraies lorsque nos phrases correspondent à un état de choses, et fausses lorsqu'elles n'y correspondent pas. La direction d'ajustement qui nous intéresse, dans ce cas, va des mots au monde. Mais parfois c'est la relation inverse qui nous intéresse, celle dans laquelle nous faisons en sorte, par nos actions, que le monde s'ajuste à nos paroles, à nos représentations. Dans ce deuxième cas, la direction d'ajustement va du monde aux mots, ou aux représentations. Nous pourrions dire, en adoptant le vocabulaire kantien (auquel Austin fait allusion, d'ailleurs), que nous nous faisons, alors, « causes » de la réalité de nos représentations. C'est le verbe qui commande, oriente, détermine.

Cependant, cette position causale, disant que nos mots visent à transformer ou à réaliser des états de choses, ne peut pas être attribuée raisonnablement à Austin, car elle supposerait qu'il instaure un écart temporel entre l'antécédent, le dire, le conséquent, et le faire. Or le célèbre slogan « Quand dire, c'est faire » établit une *simultanéité* entre dire et faire et non une *succession*.

Il n'est pas très difficile de trouver des remarques d'Austin très précises sur ce sujet qui devraient nous empêcher de donner une signification étroitement causale à la relation qu'il entend établir entre les mots et le monde.

Même lorsqu'il nous invite à distinguer l'aspect dit perlocutoire de l'énoncé (celui qui peut s'exprimer dans un verbe causal tel que persuader), de l'aspect illocutoire (qui peut s'exprimer dans un verbe d'initiative tel qu'avertir), Austin modère le sens causal : « Le sens selon lequel dire quelque chose produit des effets sur d'autres personnes ou cause un effet est foncièrement différent du sens de causer lorsqu'il s'agit d'une causalité physique, par pression, etc. La première causalité doit opérer à travers les

1. Anscombe G.E.M., *op. cit.*, 1957. Cf. aussi Searle J., *op. cit.*, 1985, pour l'expression « direction d'ajustement ».

conventions du langage et relève de l'influence qu'exerce une personne sur une autre : ce qui est sans doute le sens primitif de « causer »¹. Ainsi, même la prétendue « causalité » perlocutoire passe par la reconnaissance, la compréhension, les dispositions de l'auditeur. Il n'y a pas de causalité directe par les mots (il y a au plus « influence »).

Cependant, on peut donner un certain contenu causal à l'aspect perlocutoire, par opposition à l'illocutoire, si nous nous référons à la caractérisation de la relation causale que nous avons proposée au début de cet exposé. Ce qu'Austin appelle perlocutoire peut être dit causal dans la mesure où la relation qui est postulée n'est pas nécessaire ou connaissable *a priori*. En essayant d'avertir quelqu'un, il peut se faire que je le vexe ou que je l'irrite ou encore que je l'effraie. En revanche, il ne peut pas se faire que je n'avertisse pas quelqu'un en l'avertissant sérieusement. Il ne s'agit pas d'une conséquence matérielle, ni d'une conséquence réflexive, ou purement interne au langage. Car, nous l'avons vu, Austin n'admet pas la réduction de l'illocutoire au réflexif. Nous devons donc chercher, comme Austin nous y invite, me semble-t-il, à caractériser cette relation en des termes qui ne soient ni causaux, ni réflexifs.

Austin propose le terme *conventionnel*. On pourrait dire que l'avertissement est une conséquence conventionnelle du fait que je dis telle ou telle phrase. Toutefois, le terme conventionnel est très ambigu chez Austin. Tantôt il semble signifier la possibilité purement linguistique d'utiliser un verbe indiquant l'acte accompli en préfixe à une phrase (je te conseille, je t'avertis, je te suggère), par opposition à l'impossibilité de réaliser ce préfixe pour les verbes causaux (je te convaincs, je te persuade, je te vexe). Tantôt « conventionnel » semble signifier tout autre chose, qui me paraît beaucoup plus stimulant.

À un certain endroit, Austin indique assez clairement ce qu'il entend par une conséquence qui ne serait ni causale, ni réflexive. Que dit-il ? « Supposons maintenant que notre acte ait été effectué : tout s'est passé normalement, et aussi, si vous le voulez, sincèrement. Dans ce cas, l'énoncé performatif a l'habitude de "prendre effet". Par cela, nous ne voulons pas dire que tel ou tel

1. Austin J.L., *op. cit.*, 1970, p. 122 (note).

événement futur est ou sera produit comme effet de cet acte comme cause. Nous voulons plutôt dire que par suite de l'accomplissement de cet acte, tel ou tel événement futur, s'il arrive, sera en règle, et que tel ou tel autre événement, s'il arrive, ne sera pas *en règle*. Si j'ai dit "je le promets", je ne serai pas en règle si je manque à ma parole ; si j'ai dit "je vous souhaite la bienvenue", je ne serai pas en règle si je me mets à vous traiter en ennemi ou en intrus¹ ».

Ainsi, Austin nous met en garde contre une interprétation étroitement causale de « prendre effet ». Dans le cas des promesses, il n'y a pas un antécédent (la phrase, « Je le promets ») qui devrait nous servir à prédire le conséquent (il fera probablement ce qu'il a promis). En fait, il y a une inversion de la directionnalité. C'est seulement l'accomplissement ou le non-accomplissement de l'acte qui donne une valeur rétrospective à la promesse. Et, pas plus qu'on ne dirait d'une contradiction logique qu'elle est un événement qui s'ajoute à une affirmation et à sa négation, on ne devrait dire (ou penser) que ce qui est fait ou qui n'est pas fait relativement à la promesse s'ajoute à la promesse. Lorsqu'au cours d'une conversation j'affirme que « la terre est plate », puis lors d'un autre tour de parole, j'affirme que « la terre n'est pas plate », je me contredis. Mais c'est ma seconde affirmation ou la négation de la première qui engendre la contradiction. On pourrait parler d'une sorte de causalité rétrospective, d'engendrement de l'antécédent par le conséquent.

Mais il faut utiliser le verbe « engendrer » avec toutes les précautions possibles dans ce cas. Car, en réalité, la contradiction n'est pas un événement supplémentaire qui s'ajouterait pour ainsi dire à mon affirmation et à la négation de cette affirmation.

Ce qu'Austin nous suggère à propos de la promesse est proche, mais pas tout à fait identique. Lorsque je m'engage à faire quelque chose, ma promesse ne conditionne pas causalement la réalisation de l'état de choses promis. Mais si je ne fais pas ce que j'ai promis de faire, il y a un trouble rétrospectif : j'engendre un abus, par infraction aux conventions, comme on peut engendrer une contradiction par infraction aux règles logiques. L'abus est donc une conséquence qui n'est ni logique ni causale, bien qu'elle ressemble plutôt à une conséquence logique.

1. Austin J.L., *op. cit.*, 1962 a, p. 272.

L'idée qu'il existe des conséquences qui ne sont ni logiques ni causales a été défendue par Goldman. Il appelle d'ailleurs ces conséquences : *engendrement conventionnels*¹. Ainsi, lorsque Ravaillac poignarde Henri IV, son acte est, dans ces circonstances, un crime. Mais le fait que cet acte devienne un crime n'est pas un événement qui s'ajoute à l'acte, bien qu'en un certain sens, il en soit une conséquence, c'est-à-dire quelque chose de différent de l'acte lui-même. Pour Goldman en effet, « donner un coup de poignard » n'est pas identique à « commettre un crime », dans la mesure (au moins) où ces deux expressions ne peuvent pas être substituées dans tous les contextes (les coups de poignard qui ne sont pas des crimes ne manquent pas, en temps de guerre par exemple). Si on veut respecter l'idée de conséquence pour parler du crime sans utiliser un vocabulaire causal inapproprié, on dira que le coup de poignard de Ravaillac a engendré conventionnellement son crime. La directionnalité n'est pas respectée bien sûr, car il y a simultanéité, superposition et non succession. Il n'y a pas d'indépendance logique non plus, car le fait que l'acte de Ravaillac soit un crime ne nous est pas appris par l'observation d'autres circonstances similaires. Ce n'est pas l'expérience ou l'observation de régularités qui nous enseignent l'existence de cette relation, mais la lecture des textes de lois enregistrant des décrets humains. Évalués selon nos critères, ni l'effet de la promesse ni la transformation d'un acte en crime ne sont des conséquences logiques ou causales, bien que nous ayons intérêt à les envisager comme des conséquences.

Pour conclure, je voudrais faire une réserve. L'importance de ces relations d'engendrement conventionnel ne devrait pas nous contraindre à abandonner entièrement l'idée d'engendrement causal ou celle de réflexivité.

Cependant, même si nous choisissons d'intégrer ces trois dimensions, conventionnalité, réflexivité et causalité, dans nos descriptions de l'action humaine, nous devons éviter de donner un sens trop fort à la causalité et une place exagérée à la réflexivité. C'est ce qu'Austin avait suggéré, ou, du moins, c'est ce que j'ai essayé de lui faire dire.

1. Goldman A.I., *A Theory of Human Action*, Englewood Cliffs, Prentice Hall, 1970. Cf. le compte rendu et l'analyse de J.L. Petit « La Théorie de l'action : comment la faire ? » dans *La sémantique de l'action*, P. Ricoeur (éd.), Éd. du C.N.R.S., 1977, p. 189-222.

4. NORMES, RÈGLES, LOIS, ETC.

Bien que l'idée de norme soit l'une de celles qui ont le plus servi dans l'explication du comportement humain, elle est loin d'avoir reçu le traitement conceptuel qu'elle méritait.

C'est, du moins, ce qu'affirmait Jack P. Gibbs dans un essai publié en 1965¹ qui reste un texte de référence², la situation n'ayant guère évolué depuis, principalement parce qu'on a cherché à se débarrasser des normes, à leur substituer des notions moins compromettantes (conventions, règles, cadres de référence, modèles, etc.) au lieu d'affronter directement la difficulté³. Mais on ne se débarrasse pas si facilement de ce genre d'idée très générale.

Lorsque les sociologues se sont découragés, les philosophes, les juristes, les psychologues les ont naturellement remplacés. On a continué à s'intéresser aux normes en philosophie morale, en éthique appliquée, dans la philosophie du droit, du langage, de l'esprit, et même dans les sciences dites cognitives⁴.

1. Gibbs, J.P., « Norms. The Problem of Definition and Classification », *The American Journal of Sociology*, 70, 1965, p. 586-593.

2. Cf. l'article « Norme (sociale) » de François Chazel, dans *l'Encyclopédie universelle de philosophie*, Paris, PUF, 1991, p. 1768-1769.

3. Le diagnostic récent de François-André Isambert rejoint en grande partie celui de Gibbs : « Durkheim et la sociologie des normes », dans Chazel F. et Commaille J. eds, *Normes juridiques et régulation sociale*, Paris, Librairie générale de Droit et de Jurisprudence, 1991, p. 51-64.

4. Le symposium organisé par la revue *Ethics*, 100, 1990, « On Norms in Moral and Social Theory » témoigne bien de cette effervescence. On peut y lire, entre autres, les contributions de philosophes qui, tout en étant plus ou moins « analytiques », défendent des points de vue fort différents sur les normes. Cf. Philip Pettit, « Virtus Normativa, Rational Choice Perspectives », p. 725-755 ; Alain Gibbard, « Norms, Discussion and Ritual : Evolutionary Puzzles », p. 787-802 ; Christina Bicchieri, « Norms of Cooperation », p. 838-861 ; Jon Elster, « Norms of Revenge », p. 862-885. Dans toutes ces discussions, on retrouve, à des degrés divers,

Finalement, le sujet se retrouve dans l'agenda des sociologues, tantôt par simple contagion, tantôt pour des raisons plus respectables¹. Il me semble qu'on peut à nouveau s'engager dans un débat sur les normes en général, sans donner l'impression de vouloir naïvement ressusciter de vieilles lunes.

1

On peut distinguer, à première vue, trois sens au moins des termes « norme » ou « normatif »².

... l'influence de David Lewis, dont le livre *Convention*, Oxford, Basil Blackwell, 1969, est incontestablement à l'origine du renouveau de la réflexion sur les normes dans la philosophie analytique. En dehors de la philosophie analytique, l'idée de norme est également au centre de nombreux débats, cf., entre autre, Bubner R., « Norme et histoire », dans « Vingt ans de pensée allemande », *Critique*, 413, 1981, p. 929-949. Dans cet essai, Rudiger Bubner donne de précieuses indications sur la réflexion sur les normes dans la philosophie du droit.

1. Il y en a quatre (au moins), partiellement liées entre elles.

a. Le renouveau de l'intérêt pour la sociologie morale de Durkheim.

b. La réhabilitation ou la réappropriation des théories de Parsons pour contenir les ambitions quelque peu envahissantes des partisans de la généralisation du modèle de l'agent rationnel (cf. Jeffrey Alexander, « The Parsons Revival in Germany », *Sociological Theory*, 2, 1984, p. 394-412). Sur le débat agent rationnel-agent normatif, voir le panorama d'Abell, précis et concis : Peter Abell, « Homo Sociologicus : Do we need Him/Her ? », *Sociological Theory*, 9, 2, 1991, p. 195-198.

c. La réflexion autour d'Habermas, qui a redonné une sorte de dignité au thème des normes en substituant aux théories sociologiques traditionnelles *relativistes* et *émotivistes* une théorie *universaliste* et *cognitiviste* de la norme (cf., en particulier, Paul Ladrière, « Le conflit entre rationalité cognitive-instrumentale et rationalité morale-pratique dans la sociologie du droit de Max Weber », dans Chazel F. et Commaille J., éd., *op. cit.*, p.265-271).

d. Le débat autour des « conventions », dont on peut se faire une bonne idée en lisant le dossier présenté par Louis Quéré et Alain Rallet, « Conventions », *Réseaux*, CNET, 1993. Un numéro de *Sociological Theory*, 9, 2, 1991, donne un bon aperçu de l'ensemble de ces directions de recherche.

2. Cf. la plupart des bons dictionnaires, spécialisés ou non, ainsi que Engel P., « Trois formes de normativité », dans *Lire Davidson, interprétation et*

1) Prescriptif

Les énoncés contenant des expressions d'obligation, d'interdiction, de permission peuvent être dits normatifs. Certains sont formulés sur le mode impératif (« Venge-toi », « Ne porte pas la main sur ton père », etc.). Mais la présence d'indicateurs déontiques tels que « il faut », « il est permis, interdit, obligatoire », « tu peux », « tu dois », dans des phrases indicatives ordinaires, suffit à donner un caractère normatif à un énoncé.

2) Appréciatif

Les énoncés contenant des expressions d'évaluation peuvent être dits normatifs. On les reconnaît à la présence de prédicats axiologiques généraux tels que « bien », « mal », « beau », « laid », ou spécifiques tels que « courageux », « lâche », « mesquin », « subtil », « harmonieux », « rationnel », « irrationnel ». On peut, au moyen de ces prédicats, former des jugements moraux tels que : « Ce type est une canaille » ou esthétiques tels que « Cette sonate est sublime », et classer ou hiérarchiser des objets, actions, états (produits de qualité « supérieure », « inférieure », de « premier choix », de « bas de gamme », travail « insuffisant », « méritoire », « excellent », performances « exceptionnelles », « médiocres », « catastrophiques »). En règle générale, ces prédicats évaluatifs ou appréciatifs sont dits « attributifs » ou « relatifs » (petite taille pour un éléphant mais grande pour une souris, etc.)¹. Mais on peut admettre qu'il existe aussi des usages « absolus » (l'usage « moral », par exemple : « torturer est mal », absolument). On peut avoir l'impression que les sens évaluatif ou appréciatif se distinguent du sens prescriptif du fait que l'évaluation porte sur ce qui est, sur des états de choses *préexistant* au jugement, alors

... *et holisme*, Combas, L'éclat, 1994, p. 205-225 ; Isambert F.-A., « Durkheim et la sociologie des normes », *op. cit.*, 1991 ; Kolakowski L., « Normes qui commandent et normes qui décrivent », dans *Normes et déviations*, Neuchâtel, La Baconnière, 1987, p. 15-47 ; Livet P., « Normes et faits », *Encyclopédie universelle de philosophie*, Paris, PUF, 1991, p. 124-132 ; Von Wright G.H., *Norm and Action*, Routledge & Kegan Paul, 1963.

1. Williams B., *Morality, An Introduction to Ethics*, Cambridge, Cambridge University Press, 1972, p. 52-61.

que la prescription porte sur ce qui n'est pas encore, sur ce qui ne préexiste pas au jugement. Cependant, on peut envisager la possibilité d'évaluations constitutives, si on peut dire. On peut avoir des raisons de penser que, lorsque nous percevons un objet comme une œuvre d'art ou un mouvement corporel comme une action, nous instituons quelque chose par l'évaluation : l'œuvre d'art ou l'action ne préexisteraient pas à nos évaluations esthétiques ou morales¹.

3) Quantitatif ou descriptif

Les énoncés qui font référence à des états moyens, habituels, fréquents, ordinaires, répandus d'objets, actions, événements peuvent être dits normatifs. Ces états moyens, fréquents, habituels peuvent servir de critères d'évaluation. Mais ce n'est pas toujours le cas et c'est pourquoi on doit distinguer le sens évaluatif ou appréciatif du sens purement quantitatif du normatif. Les énoncés qui font référence à ces états moyens, fréquents, habituels n'ont pas de caractéristiques linguistiques aussi spécifiques que les indicateurs déontiques, le mode impératif pour les prescriptions ou les prédicats axiologiques pour les évaluations. Ils sont reconnaissables néanmoins à certaines formes de *quantification vague* : « La plupart » des Français aiment le fromage ; « Les » Siciliens ont l'habitude de se venger ; « Les » courgettes sont vertes, etc. Toutefois cette propriété ne suffit pas à les distinguer des énoncés descriptifs ordinaires et c'est pourquoi on utilise indifféremment les termes quantitatifs ou descriptifs pour cette troisième variété d'énoncé normatif.

À la suite de Frege, différents sémanticiens nous recommandent de distinguer le sens d'un énoncé tel qu'il apparaît ou tel qu'il s'exprime dans sa forme linguistique, de sa force ou de l'acte qui est accompli au moyen de son énonciation. En effet un énoncé peut avoir un sens linguistique descriptif (ou une forme descriptive) et exprimer, néanmoins, une obligation ou une interdiction

1. On pourrait dire que, lorsque le philosophe Donald Davidson parle de « normativité », il pense à quelque chose qui ressemble à une « évaluation constitutive ». C'est le fait que nous avons décidé de voir certains gestes et certaines croyances sous l'angle de leur cohérence qui en fait, pour ainsi dire, des actions. Cf. Davidson D., *op. cit.*, 1993.

(exemples : « Les pâtes sont prêtes » ; « Le poisson rouge ne mange pas de cacahuètes », etc.). Si on veut tenir compte de ce fait sémantique, il convient donc de préciser que, pour être authentiquement normatifs, ces énoncés doivent, dans le premier cas, exprimer la force d'une obligation, interdiction, autorisation, etc. ; dans le second, l'existence d'un modèle, étalon, critère d'évaluation, et dans le troisième l'existence d'états moyens, fréquents, habituels.

Sans approfondir la question sémantique (sur laquelle je reviendrai plus loin), il pourrait être utile de garder à l'esprit le fait que la distinction entre le sens et la force des énoncés, de même que la différence, en partie dérivée, entre l'aspect purement linguistique et l'aspect institutionnel de la norme permettent de justifier l'idée que les énoncés normatifs sont susceptibles d'être réputés vrais ou faux.

La dichotomie traditionnelle du fait et de la valeur repose sur une thèse aujourd'hui contestée disant qu'à la différence des énoncés descriptifs, les énoncés dits normatifs n'auraient pas de conditions de vérité. Autrement dit, d'après cette thèse, il serait absurde de se demander s'il est vrai que « mentir est mal », mais on pourrait exprimer sa répulsion, son indignation à l'égard du mensonge. Bref, ce genre d'énoncé normatif ne ferait qu'exprimer les sentiments de la personne qui juge et ne pourrait en aucun cas être pris pour des descriptions de propriétés réelles, intrinsèques, indépendantes de nos sentiments, de l'institution jugée. Telle est, du moins, l'opinion des émotivistes. Mais les cognitivistes ont estimé, contre les émotivistes¹, que les énoncés normatifs pourraient être, au même titre que les énoncés descriptifs ordinaires, susceptibles d'être vrais ou faux et qu'ils auraient même, de par leur forme, une prétention à la vérité ou à l'objectivité (c'est particulièrement évident dans le cas des énoncés évaluatifs). Dans ce débat, la possibilité d'opposer le sens linguistique d'un énoncé et sa force a joué un rôle crucial et trop souvent méconnu².

1. Sur le débat émotivisme-cognitivisme, voir Monique Canto-Sperber, *La philosophie morale britannique*, Paris, PUF, 1994.

2. Comme nous le verrons au chapitre consacré au réalisme et au relativisme, certains arguments permettent d'étayer l'idée qu'au plan linguistique, une phrase telle que : « Mentir est mal » peut être jugée aussi vraie

Toutefois, la question sémantique est loin d'être réglée, et les observations suivantes s'adressent à ceux qui estiment que les énoncés dits normatifs ont des conditions de vérité aussi bien qu'à ceux qui continuent d'exclure cette possibilité.

2

Hare a essayé de montrer que les sens évaluatif et prescriptif des énoncés normatifs pouvaient être confondus¹. À première vue, il est assez difficile de lui donner tort, puisqu'on a, semble-t-il, tendance à recommander ce qu'on juge bon. Plus précisément, il semble qu'on ne puisse pas dire d'une chose qu'elle est bonne si on ne la recommande pas. Mais il existe des choses que nous apprécions ou que nous jugeons bonnes, mais dont nous savons que l'obtention ou la réalisation sont manifestement indépendantes de notre volonté, c'est-à-dire que nous ne pouvons pas les obtenir ou les réaliser sur ordre, prescription, recommandation. Nous pouvons apprécier la beauté, le naturel, la spontanéité, la foi, l'intelligence sans qu'il nous soit possible de les prescrire ou de les recommander, parce que nous savons très bien qu'avoir ces qualités n'est précisément pas une affaire de volonté, mais, au mieux, un effet indirect d'actions qui ne les visent pas, un effet

... (ou aussi fausse) que « Deux et deux font quatre », « L'argile est imperméable », etc. C'est que, si l'on s'en tient au sens linguistique de l'énoncé, et dans la mesure où l'on identifie ce sens aux conditions de vérité de l'énoncé, la question de savoir comment on prouve cette vérité, quels sont les moyens effectifs que nous utilisons pour la démontrer n'a aucune importance (on peut envisager différentes possibilités extralinguistiques : correspondance à la réalité, conformité à des principes, cohérence interne, etc.) En fait, c'est seulement lorsque nous nous intéressons à l'*usage* qui est fait de ces phrases que nous pouvons identifier des différences entre le prescriptif, l'évaluatif ou l'appréciatif, et le descriptif, du point de vue de leur rapport à la vérité. Et il n'est pas exclu qu'on admette alors qu'il existe différentes formes d'objectivité, différentes manières de prouver la vérité d'un énoncé, la correspondance à la réalité n'étant pas le seul critère d'évaluation disponible. Williams B., *Morality*, op. cit., 1972 ; Horwich P., *Truth*, Oxford, Basil Blackwell, 1990.

1. Hare R.M., *The Language of Morals*, Oxford, Clarendon Press, 1952.

non voulu¹ (je peux estimer qu'il est *bon* d'avoir du talent mais je ne peux pas *ordonner* ou *recommander* à quelqu'un d'avoir du talent).

D'autre part, les tentatives de ruiner la distinction entre les deux premiers sens, évaluatif et prescriptif, et le troisième, quantitatif, c'est-à-dire au fond descriptif, n'ont pas manqué. Il existe d'assez bons arguments pour soutenir qu'il serait vain d'essayer de séparer les aspects évaluatifs des aspects descriptifs d'énoncés tels que : « La Tour Eiffel est harmonieuse » ou « Gustave Eiffel est courageux ». Par ailleurs, chaque fois qu'on fait référence aux « besoins » d'un organisme ou d'une machine, il n'est pas très facile de distinguer les sens évaluatif et prescriptif du sens descriptif (exemples : « Il faut remplir le réservoir d'essence » ; « Il ne faut pas trop arroser cette plante », etc.)². Mais, comme Hilary Putnam l'a très justement souligné, ces arguments n'ont pas réussi à ébranler cette véritable institution culturelle qu'est, en réalité, la distinction du normatif et du descriptif³. Dans les tribunaux, dans les discussions et les récits ordinaires, on continue d'exiger une séparation nette entre ce qui est de l'ordre du fait et de l'ordre de la valeur, et les théories sémantiques sophistiquées qui rejettent cette distinction sont encore (pour cette raison probablement) accueillies avec le plus grand scepticisme. Tout ce qu'on peut dire à présent, c'est que nos façons habituelles de défendre l'opposition entre jugements de fait et jugements de valeur sont fragiles, et qu'il est nécessaire de trouver de meilleurs arguments pour les justifier. Mais nous n'en sommes pas encore au point où il faudrait renoncer définitivement à cette opposition. D'ailleurs, certains obstacles à la formulation d'une théorie cohérente des normes ont manifestement pour origine la volonté d'ignorer l'opposition traditionnelle entre

1. Elster J., *Le Laboureur et ses enfants* (1983-1984), trad. A. Gerschenfeld, Paris, Minuit, 1986, p. 17-90.

2. Williams B., (1985), *L'éthique et les limites de la philosophie*, trad. Marie-Anne Lescourret, Paris, Gallimard, 1990, p. 141-143. Anscombe G.E.M., « Modern Moral Philosophy », *Ethics, Religion and Politics, Collected Philosophical Papers*, III, Oxford, Basil Blackwell, 1981, p. 32.

3. Putnam H., *Raison, vérité et histoire* (1981), trad. A. Gerschenfeld, Paris, Minuit, 1984, p. 145-168.

le prescriptif, l'évaluatif et le descriptif. Mais la principale provient, me semble-t-il, du fait que la norme ne peut, en définitive, être réduite à aucune de ces caractérisations particulières (prescriptive, évaluative, descriptive) et qu'elle doit, d'une façon ou d'une autre, les intégrer toutes. C'est pourquoi on a pris l'habitude, en sociologie, par exemple, de proposer des définitions *pluridimensionnelles* de la norme telles que : mode de comportement prescrit, apprécié et répandu. La question de savoir si un comportement est répandu *parce qu'il* est prescrit ou apprécié ou s'il est prescrit et apprécié *parce qu'il* est répandu est d'un autre ordre, bien évidemment, car ce n'est pas une question de définition.

Cependant, en procédant de cette façon, on est loin d'avoir réussi à identifier un ensemble de conditions nécessaires et suffisantes qui permettraient de construire une sorte de définition satisfaisante de la norme, c'est-à-dire pouvant faire l'objet d'un certain accord. C'est, du moins, la conclusion qu'on peut tirer des observations de Gibbs, pour qui la discussion sur les normes souffre de trois déficits¹ :

- 1) l'absence d'accord sur les définitions générales
- 2) l'absence d'accord sur les typologies
- 3) l'absence d'accord sur ce qu'il faut considérer comme des propriétés non contingentes des normes (vraies, par définition, vraies pour tous les types de normes), et ce qu'on peut considérer comme des propriétés contingentes (vraies pour certains types de normes seulement).

Gibbs s'appuie sur un échantillon de définitions puisé dans un ensemble extrêmement peuplé. Mais, au vu de l'échantillon, on pourrait objecter qu'il a pour objet spécifique les normes sociales par opposition, disons, aux normes techniques, morales, juridiques, et qu'il n'est pas étonnant, par conséquent, qu'il ne permette pas de construire une définition ayant la généralité requise.

Toutefois, étant donné que ce qui est en question c'est, précisément, la validité de ce genre de classification ou de typologie, l'objection ne tient pas. En réalité, il vaut mieux faire comme Gibbs, partir de quelques définitions existantes, sociologiques

1. Gibbs J.P., *op. cit.*, 1965.

par exemple (pourquoi pas ?), et se poser ensuite la question de leur généralité. Je vais m'inspirer de cet excellent exemple en partant du groupe de définitions qu'il a sélectionné, mais en choisissant parmi elles, celles qui sont les plus opposées, pour bien marquer la difficulté.

— Une norme est un énoncé émis par un certain nombre de membres d'un groupe, pas nécessairement par tous, disant que les membres *doivent* se comporter de telle ou telle façon dans telle ou telle circonstance¹.

— Une norme est une règle ou un critère gouvernant notre conduite dans les situations sociales où nous sommes engagés. C'est une *attente* sociale. C'est ce à quoi on s'attend que nous nous conformions, que nous le fassions ou pas².

— Une norme est un modèle abstrait dont le siège est dans l'esprit, qui trace des limites aux comportements. La norme effective est celle qui n'est pas simplement dans l'esprit, mais celle qui est aussi tenue pour une norme qui mérite d'être appliquée dans un comportement effectif, de sorte qu'on peut sentir qu'on doit s'y conformer. C'est ce sentiment qui indique qu'on « accepte » la norme³.

— Les normes sont des prescriptions ou des interdictions sanctionnées et largement acceptées, portant sur les comportements, croyances, sentiments d'autrui, c'est-à-dire sur ce que les autres doivent faire croire, sentir, etc. Les normes doivent être des prescriptions *partagées*. Elles incluent toujours des sanctions⁴.

Il n'est pas difficile de voir que cet ensemble de définitions ne nous permet pas de sélectionner un ensemble de conditions nécessaires et suffisantes, qui nous autoriserait à caractériser la norme de façon formelle, explicite, générale.

1. Homans G., *Social Behavior : Its Elementary Forms*, New York, Harcourt, Brace and Co, 1961, p. 46.

2. Bierstedt R., *The Social Order*, 2^e éd., New York, Mc Graw-Hill Book Co, 1963, p. 222.

3. Johnson H.M., *Sociology*, New York, Harcourt Brace and Co, 1960, p. 8.

4. Morris R. T., « A Typology of Norms », *American Sociological Review*, XXI, 1956, p. 610.

En effet, la confrontation de ces définitions indique que les propriétés jugées essentielles par tel auteur sont jugées accidentelles par d'autres. Et c'est bien ce qui interdit la construction de typologies convaincantes. Ainsi, selon que les sanctions sont ou non considérées comme des propriétés essentielles des normes sociales, les résultats taxonomiques sont différents. Or, il se trouve que la sanction n'est pas toujours tenue pour une propriété intrinsèque de la norme.

Il est vrai que, dans l'ensemble, ces définitions semblent mettre l'accent sur le caractère collectif, prescriptif, sanctionné de la norme. Mais un peu d'attention montre qu'*aucune* de ces propriétés n'est nécessaire ou suffisante.

Tout d'abord, certaines définitions reconnaissent qu'un certain genre de comportement peut représenter la norme, même s'il n'est pas le plus fréquent ni même le plus largement approuvé dans une population donnée. D'ailleurs, la détermination d'une sorte de seuil quantitatif poserait des problèmes insurmontables. Tel ou tel type de comportement cesse-t-il d'être une norme aussitôt qu'il n'est plus suivi ou approuvé par cinquante pour cent au moins d'une population donnée ? Mais comment distinguer ce cas de celui dans lequel il y aurait (provisoirement) un niveau « exceptionnel » de déviance ? Ne suffit-il pas, pour qu'un certain genre de conduite soit considéré comme une norme, qu'il soit soutenu, approuvé, sanctionné par ceux qui sont particulièrement bien situés, au sein d'une population donnée, dans la hiérarchie du prestige du pouvoir, des biens économiques ? Mais si on l'admet, on abandonne manifestement les critères quantitatifs.

Les définitions qualitatives insistent moins sur l'aspect collectif que sur l'aspect prescriptif de la norme. Mais, pas plus que le caractère collectif, répandu, fréquent de la norme, ce caractère prescriptif ou impératif ne peut être tenu pour une propriété essentielle ou un attribut nécessaire de la norme. Aussi curieux que cela puisse paraître, il est possible de définir la norme sans faire intervenir l'idée de prescription. C'est bien ce que font ceux qui insistent sur l'aspect prédictif de la norme.

Dans nos relations avec autrui, nous avons des attentes, nous anticipons des comportements. Nous pensons aussi que les autres ont des attentes et des anticipations du même genre à

notre égard. Nous réglons nos conduites sur ces attentes et sur ces anticipations réciproques, sans exclure la possibilité de manipulations auxquelles elles peuvent donner lieu. Ce sont ces attentes et ces anticipations qui nous font servir du café à la fin du repas plutôt qu'au début, ou qui font que nous nous attendons à en recevoir à la fin plutôt qu'au début. Ce qui justifie, à nos yeux, ces attentes réciproques, c'est notre croyance en l'existence de certaines formes de régularité dans les conduites, qu'on peut appeler des normes. La question de savoir si l'existence de ces régularités est le produit d'une conformité consciente ou inconsciente à un ensemble de prescriptions n'est pas la plus importante. On peut envisager différentes façons d'expliquer ces régularités parmi lesquelles celle qui postule la conformité aux prescriptions ne sera peut-être pas la plus convaincante.

Un autre moyen de caractériser la norme sans faire intervenir l'idée de prescription consiste à mettre l'accent sur l'aspect attractif de la norme. Une règle ne devient une norme que si nous l'approuvons, si nous pensons qu'elle est bonne et non si elle se présente à nous comme une obligation purement extérieure. Il existe différentes stratégies durkheimiennes pour faire coïncider le plaisir et le devoir. Mais ceux qui ont choisi, par parti pris philosophique ou autre, d'opposer les théories attractives et les théories impératives du bien, rangeront certainement le comportement normatif du côté de l'attractif, plutôt que de l'impératif.

Bref, pour les prédictivistes et les attractivistes, le domaine de la norme ne coïncide pas avec celui des règles, devoirs, prescriptions, lois, conventions, obligations, interdictions, etc. Ils n'ont probablement pas tout à fait tort puisque nous choisissons souvent de ne pas confondre les normes et les règles, prescriptions, lois, conventions, devoirs. On peut parler, par exemple, des règles du jeu d'échecs, mais ces règles ne sont pas tenues pour des normes ; de même on parle des lois d'une puissance militaire occupante, des règlements d'une prison, ou des devoirs envers Dieu, sans penser qu'on pourrait substituer norme à loi, règlement, devoir dans ces cas. Il semble qu'on ait pris l'habitude de dire des lois, règles, prescriptions, conventions, devoirs, qu'ils sont des normes, si et seulement si, ils sont approuvés, appréciés, subjectivement désirés.

Du fait que la norme ne se confond pas exactement avec les lois, prescriptions, règles, devoirs, il semble suivre qu'elle n'est pas nécessairement ou intrinsèquement liée à l'idée de sanction. Il est évident qu'il ne suffit pas qu'il y ait sanction pour qu'il y ait norme, puisqu'on peut envisager la possibilité qu'existent des règles sanctionnées qui ne soient pas des normes. Mais il peut sembler aussi à présent que l'existence des normes n'est pas nécessairement garantie par celle des sanctions.

Certaines analyses suggèrent qu'il convient de séparer les sanctions internes (honte, culpabilité, embarras, mais aussi satisfaction, orgueil, fierté) des sanctions externes (sarcasmes, rires, mépris, châtiments physiques, mais aussi félicitations, louanges, récompenses matérielles). Et elles laissent entendre, de plus, que les sanctions internes ne méritent pas à proprement parler le nom de sanction. C'est particulièrement évident dans le cas des sanctions positives internes. La satisfaction du devoir accompli devrait-elle être considérée comme une sanction ? Ce serait vraiment donner un sens trop large à ce terme, un sens que Durkheim, lui-même, pourtant très généreux en la matière, n'accepterait probablement pas.

Quoi qu'il en soit, certaines définitions ne font absolument pas allusion aux sanctions, fussent-elles positives et internes. D'après elles, il suffirait, pour que l'existence d'une norme soit avérée, de constater que certaines croyances sont largement partagées ou distribuées dans une population donnée, ou, plus exactement, dans les « esprits », les consciences individuelles, dont cette population est formée. Ce genre de définition positive de la norme (croyances largement partagées ou distribuées dans une population donnée) contredit, bien sûr, les définitions qui n'accordent aucune importance aux quantités et aux fréquences, ce qui n'arrange rien !

On peut donc dire, en résumé, que les trois attributs prescription, évaluation, distribution ne sont ni suffisants (il ne suffit pas que quelque chose soit une prescription, ou une évaluation, ou une régularité largement distribuée pour être une norme) ni nécessaires (apparemment il n'est pas nécessaire que quelque chose soit prescrit, ou apprécié, ou largement répandu pour être une norme).

À ces problèmes conceptuels, s'ajoutent des problèmes techniques qui se posent, semble-t-il, quelle que soit la définition de la norme retenue, et dont la critique des ethnométhodologues (entre autres)¹ nous a fait prendre conscience. Je les appellerai, pour ma part, problèmes de l'indétermination subjective et objective de la norme. Commençons par l'indétermination subjective.

Dans la mesure où l'idée de norme fait référence, dans la plupart des théories, à des croyances, des sentiments, des attitudes, des pensées subjectifs, autant que des règlements, lois, instructions objectifs, la question de leur identification se pose de façon aiguë.

Il n'est pas nécessaire d'être un sceptique incurable pour soutenir que nous n'avons, à l'heure actuelle, aucun moyen sérieux d'identifier ce qui se passe « dans la tête » des agents, ni même de le simuler. Ni les opinions exprimées en situation d'enquête, ni les comportements effectifs ne peuvent donner un contenu à ces croyances, pensées, attitudes (ni les opinions exprimées parce qu'elles ne correspondent pas nécessairement aux attitudes qui orientent le comportement, ni le comportement effectif parce que son absence ne signifie pas nécessairement que l'agent n'avait pas à l'esprit telle ou telle norme qu'il n'a pas réussi à rendre effective).

Quant aux typologies de normes, aux divisions en sous-ensembles tels que normes techniques, juridiques, sociales, morales, etc., elles nous confrontent aux mêmes difficultés, dans la mesure où ces classifications se font généralement sur la base des formes de sanctions : institutionnelles ou légales pour les normes juridiques ; diffuses ou informelles pour les normes sociales, etc. De ce fait, elles comportent toujours un élément subjectif difficile à identifier, un aspect statistique difficile à déterminer et un problème conceptuel impossible à éliminer (la sanction est-elle un indice de la norme ?). Passons au problème de l'indétermination objective.

1. Cicourel A. V., *La sociologie cognitive* (1972), trad. J. et M. Olson, Paris, PUF, 1979, p. 37-52.

Il serait absurde d'envisager la possibilité qu'il existe, pour toute circonstance déterminée, une loi indiquant ce qu'il convient de faire et de ne pas faire. Cependant, il n'est pas impossible d'essayer de déduire une partie, au moins, des prescriptions particulières à partir du contenu des lois générales.

Dans le cas des règles extra-légales, la situation est encore plus complexe. Il est absurde d'envisager la possibilité qu'il existe pour toute circonstance particulière une norme indiquant ce qu'il faut faire ou ne pas faire. Mais, dans le cas de normes, on n'a même pas la possibilité de s'appuyer sur quelques principes généraux ou normes générales explicites pour essayer de déduire, pour ainsi dire, les solutions particulières appropriées.

Jon Elster pense qu'on ne peut pas savoir *a priori* s'il existe une norme nous interdisant de proposer de l'argent à quelqu'un pour prendre sa place dans une file d'attente devant un cinéma, et, au cas où une tentative de ce genre échouerait, si les raisons de l'échec viendraient de ce que la norme interdisant d'acheter sa place dans une file d'attente est déduite d'une norme plus générale condamnant la corruption (on ne peut pas tout acheter), ou d'un quelconque autre principe général, auquel nous n'avons pas pensé, mais qui serait forcément aussi contestable, car ce genre de principe n'est nulle part explicité¹.

Cependant, tous ces déficits conceptuels ne devraient pas nous décourager définitivement. Du fait que nous sommes apparemment dans l'incapacité de proposer une définition formelle, explicite, générale et acceptée de la norme ou que nous n'avons aucun moyen de savoir avec précision ce qu'elles sont effectivement, dès lors qu'elles se présentent, partiellement au moins, comme des objets subjectivement et objectivement indéterminés, il ne résulte pas que la notion de norme soit inutile ou bonne à jeter.

C'est le destin de ces concepts très généraux (vérité, rationalité, réalité, bien, etc.) de rester à jamais des sortes d'idées régulatrices, dont il serait absurde de réclamer autre chose qu'une caractérisation très vague. Tout le monde sait très bien que les tentatives de « définir » la vérité, la rationalité, la réalité, le bien, aboutissent à des catastrophes intellectuelles au pire, à des

1. Elster J., *Nuts and Bolts for the Social Sciences*, Cambridge University Press, 1989, p. 115.

réductions insatisfaisantes au mieux. Mais du fait que des concepts tels que vérité, rationalité, réalité, bien ne peuvent pas être définis précisément (ou qu'il n'y a aucun sens à vouloir les définir de cette façon), on ne peut pas conclure qu'il vaut mieux s'en débarrasser. Il en va de même pour la norme.

4

On pourrait objecter toutefois que la quête d'une définition même vague de ce genre de notion est un égarement philosophique typique.

Le philosophe serait victime d'une sorte d'« illusion essentialiste ». Il aurait la naïveté de croire que ces notions pourraient avoir un sens en dehors des théories où elles prospèrent naturellement. En réalité, « norme » serait un terme *théorique* et sa signification dépendrait de certains choix conceptuels plus généraux. « Norme » ne signifie pas la même chose pour un culturaliste, un fonctionnaliste, un interactionniste, etc. Au demeurant, dans les sciences humaines actuelles, l'importance de l'idée de norme est très loin d'être unanimement reconnue. Certes, il existe des théories qui continuent de placer la norme au centre de leurs préoccupations (principalement, les différentes formes de culturalisme ou de structuro-fonctionnalisme, anciennes et « néo »). Mais les théories de l'agent rationnel destituent la norme de son rôle central traditionnel¹. Certains partisans de la théorie de l'agent rationnel vont même assez loin de ce point de vue puisqu'ils estiment qu'il faut *éliminer* la norme du vocabulaire théorique, sous prétexte que l'idée de norme est intrinsèquement liée à une conception « sursocialisée » de l'homme². Les adversaires des normes ne manquent jamais une occasion d'exprimer leur répulsion envers l'image de l'homme « conditionné ». Apparemment, ils préfèrent celle de l'« entrepreneur » dont les engagements ne valent que dans la mesure où ils satisfont ses

1. Pour l'opposition agent normatif — agent rationnel, cf. P. Abell, « Homo Sociologicus : Do We Need Him/Her ? », *op. cit.*, 1991.

2. Wrong D.H., « The Oversocialized Conception of Man in Modern Sociology », *American Sociological Review*, 2, 26, 1961, p. 183-193.

intérêts à court ou long terme. Bref, dans le contexte de la théorie de l'agent rationnel, on se passe assez bien des normes. C'est un choix théorique. Mais d'autres choix sont aussi légitimes. En réalité, là où le philosophe ne voit que de la confusion et du désaccord, il n'y a que des choix différents également justifiés.

Ce genre d'objection ressemble plus à un réflexe relativiste assez traditionnel qu'à une appréciation réelle de la difficulté. Elle semble dire en effet que les théories sont incommensurables et qu'elles ont toutes la même valeur. C'est, bien sûr, une thèse trop forte et indéfendable parce qu'elle est auto-destructrice comme toutes les thèses authentiquement relativistes (pourquoi la thèse disant que les théories sont incommensurables et ont la même valeur aurait-elle *plus* de valeur que la thèse inverse disant que les théories ne sont pas incommensurables et qu'elles n'ont pas la même valeur ?).

En général, le relativiste « théorique » est un peu plus prudent. Il est plutôt « pluraliste » qu'authentiquement relativiste. Il admet que certaines théories valent tout de même plus que d'autres, même si plusieurs théories ont approximativement la même valeur. Il admet donc qu'il n'est pas impossible de confronter les théories et de les évaluer. Selon quels critères ? Qu'est-ce qui lui permet de dire que telle théorie de la norme est en définitive meilleure qu'une autre ? S'il répond qu'il n'accepte pas d'évaluer des concepts particuliers mais seulement des théories générales (culturalisme, fonctionnalisme, interactionnisme, théorie de l'échange, etc.) la difficulté est repoussée ; elle ne disparaît pas. Quels critères vont nous aider à départager ces théories générales ? Leur capacité à nous donner la meilleure explication possible de tel ou tel phénomène ? Leur cohérence ? Leur simplicité ?

Dans les faits, ce sont souvent des arguments portant sur des concepts particuliers qui sont échangés. On dit, par exemple, que les culturalistes se font une idée erronée de ce qu'est une norme, ou que les rationalistes estiment à tort qu'une norme ne peut pas être le motif d'une action humaine, mais seulement une ressource ou une contrainte dans l'action, qui n'a pas plus d'importance que d'autres contraintes (économiques ou politiques).

Dans les faits, on traite la norme comme une sorte d'espèce naturelle que les différentes théories doivent appréhender de la meilleure façon possible (ou dont elles doivent montrer pourquoi elle est impossible à appréhender). Ce n'est probablement qu'une stratégie de recherche, mais je ne vois pas comment on pourrait l'éliminer sans se perdre dans un relativisme théorique assez stérile.

DEUXIÈME PARTIE

LA VIE INTÉRIEURE

1. NI MENTALISTE, NI BEHAVIORISTE

Nous n'avons pas fini encore de mesurer toutes les conséquences de ce qu'on peut appeler la ruine du behaviorisme dans les sciences humaines. Cette ruine a rendu à l'étude scientifique un ensemble de phénomènes dits intérieurs : les intentions, les croyances, les désirs, mais aussi les sentiments, les émotions, la conscience.

Ce sont les sciences cognitives qui ont assuré l'héritage anti-behavioriste. Mais elles ont été victimes d'une sorte de dérive « mentaliste ».

C'est une dérive qui n'était nullement impliquée dans l'engagement anti-behavioriste initial.

1

Les behavioristes avaient proscrit de leur vocabulaire un ensemble de termes dont nous nous servons habituellement pour décrire l'action humaine : pensée, croyance, idée, choix, intention, émotions, conscience, volonté, etc. Bref, tout ce qui concernait la « vie intérieure », inobservable.

Les behavioristes pensaient qu'on pouvait substituer à ces termes dénotant des entités mystérieuses des termes d'observation directe ou indirecte (tels que disposition à réagir, régularité comportementale, apprentissage, sanction, etc.).

Mais il est apparu qu'il était absurde de traiter la vie intérieure comme une « boîte noire », en particulier lorsqu'il s'agissait d'expliquer des processus aussi centraux que ceux de la maîtrise du langage. Étant donné la disparité entre la pauvreté des stimuli auxquels celui qui apprend un langage est exposé et la richesse du langage qu'il finit par maîtriser, il semble nécessaire de postuler l'existence d'une sorte de « faculté » interne intermédiaire. Bien que de nombreux aspects de cette thèse anti-behavioriste aient été contestés (en particulier la nécessité de postuler l'existence d'une faculté de langage innée et spécialisée), les prin-

cipes généraux qui la justifiaient sont restés à peu près intacts. La possibilité de simuler mécaniquement ce monde intérieur « inobservable » a contribué à les rendre plus crédibles encore. Et le vocabulaire de la « vie intérieure » s'est trouvé enrichi de termes théoriques tels que : anticipation, représentation, compétence, état computationnel, connaissance tacite, processus périphériques et centraux, modularité de l'esprit, plan, etc.

Le rejet du behaviorisme s'est généralisé et, en sociologie, il s'est appuyé sur des traditions solides : celle des sociologies compréhensive et phénoménologique, particulièrement concernées par le sens visé par l'acteur, ses projets, ses fins, ses valeurs, ses manières de reconnaître et de constituer le monde social dans l'échange intersubjectif.

De façon générale, le vocabulaire de la subjectivité et de l'intersubjectivité, celui de l'action, de la conscience, de la personne, et même de la liberté a cessé d'être suspect en sciences humaines.

C'est dans les tentatives d'élaborer une théorie systématique de la signification que la réhabilitation des notions intentionnelles (ou intérieures) a été la plus spectaculaire. Mais on ne peut pas dire que la sociologie ait été à la pointe de ses recherches sur le sens. Les sociologues les plus enclins à utiliser le vocabulaire de la signification (phénoménologues, « wébériens », etc.) se sont contentés de généralités. Ils se sont souvent satisfaits de concepts aussi vagues que ceux du « vécu », de « sens visé », de « fins » de l'acteur, en laissant de côté, par exemple, les problèmes infiniment complexes de la compréhension du sens des phrases dans les échanges quotidiens (à la notable exception, il faut le dire, de ceux qu'on appelle ethnométhodologues).

C'est à la lisière de la sociologie, dans les études de pragmatique linguistique, que les contributions à la théorie systématique de la signification ont été les plus importantes. Mais ce sont aussi celles qui, progressivement intégrées dans les sciences cognitives, ont fait le plus pour détacher l'étude du sens de la sociologie.

Les études de pragmatique ont montré qu'il convenait de séparer la signification étroitement linguistique de l'énoncé (ce qui est « dit » littéralement) du message véhiculé (ce que le locuteur a voulu dire). Nous saisissons un énoncé littéralement si nous comprenons ses « conditions de vérité », c'est-à-dire, si nous arrivons à nous faire une représentation de l'état de choses qui serait réalisé

si elle était vraie. Cette saisie est, en un certain sens, involontaire, automatique — bien que nous soyons dans l'obligation d'enrichir la phrase entendue, en fixant ses références et en éliminant les ambiguïtés, car la « proposition » exprimée par le locuteur (ce qui, dans ce qu'il dit, peut être réputé vrai ou faux) est toujours sous-déterminée par sa forme linguistique. Cependant l'énoncé littéral ainsi compris ne nous indique pas comment il doit être pris, quelle est sa « force » (ou son « mode »). La même phrase « Tu seras sage » peut être une prédiction, un ordre, un conseil, etc. La « force » de l'énoncé doit être séparée de la signification littérale étroite mais elle contribue, bien sûr, au sens général de l'énoncé.

Après une période de classement débridé d'énoncés selon leur « force » (constatifs, performatifs, etc.), il est apparu qu'il valait mieux parler de niveaux de sens¹. Au premier étage, si on peut dire : la signification linguistique étroite, constante, conventionnelle donnée par les conditions de vérité. Au second, l'interprétation des présupposés, sous-entendus, implications, etc., par des procédures non conventionnelles (hypothèses d'interprétation, sélection par pertinence...). L'idée des niveaux de sens s'applique particulièrement bien aux cas de l'ironie, de la métaphore, de l'insinuation et ainsi de suite.

Peu à peu, ces phénomènes, jugés parasites dans les théories behavioristes ou positivistes de la signification ont pris la place de faits cruciaux. C'est toute la théorie du sens qui a été reconstruite, pour ainsi dire, à partir d'eux : le sens littéral devient un cas particulier de la métaphore, l'explicite, un cas particulier du sous-entendu, etc. La séparation même des niveaux de sens a été contestée, le niveau étroitement linguistique, conventionnel, automatique, étant pratiquement éliminé.

En déplaçant l'accent de ce qui est dit vers les intentions, les vouloir-dire, ces mêmes philosophes et linguistes en sont venus à

1. Les textes d'introduction à ce débat de François Recanati sont tous très utiles, cf. entre autres, « Communication et cognition ; Dan Sperber et Deirdre Wilson, La pertinence : communication et cognition », *L'Âge de la Science*, 4, (*Philosophie de la logique et philosophie du langage*, I), Paris, O. Jacob, 1991, p. 231-249 ; « Paul Grice et la philosophie du langage ordinaire », *L'Âge de la Science*, 5, (*Philosophie de la logique et philosophie du langage*, II), Paris, O. Jacob, 1993, p. 17-22.

accorder une importance de plus en plus considérable à la pensée (par opposition au langage), aux états mentaux en général, en invitant les psychologues cognitifs à participer à leurs recherches.

Cette tendance est devenue plus marquée encore lorsque, délaissant provisoirement les problèmes de l'ironie et des sous-entendus, ces philosophes et linguistes se sont attaqués à d'autres phénomènes énigmatiques : la logique des termes indexicaux (démonstratifs tels que « ceci » ou « cela » ou égocentrés tel que « je », « tu », « maintenant », « hier », etc.), ou celle des grandes catégories grammaticales de transitivité ou d'agentivité.

Là encore, il leur est apparu qu'aucune explication satisfaisante du fonctionnement de ces expressions ne saurait être donnée si on renonçait à faire appel aux notions de conscience de soi (pour les termes égocentrés), de perception directe (pour les démonstratifs), de spatialisation ou d'agentivité (pour expliquer notre compréhension de ce qui est exprimé par ce qu'on appelle verbe transitif, catégorie trop générale qui recouvre l'action et la passion, le mouvement et le changement d'état, etc.).

Bref, la psychologie est venue s'insinuer au cœur même de ce que nous appelons un langage, ou, plus exactement, il est devenu évident, pour de nombreux philosophes, que la compréhension du phénomène de la signification demeurerait à jamais hors de notre portée, si certaines notions tombées en désuétude (telles que conscience, intention, etc.) n'étaient pas réhabilitées. Il est difficile, au point où nous en sommes, de nier l'importance de ce réseau de concepts mentaux dans l'analyse du sens.

Le projet (radicalement anti-cartésien, comme on l'a dit) d'éliminer l'esprit, de mettre un terme à nos élucubrations sur l'âme, les « fantômes » de la vie intérieure, de ranger enfin et pour toujours les concepts mentaux ou intentionnels (volitions, pensées, croyances, désirs, sentiments, etc.) dans le musée du folklore psychologique est très loin d'avoir abouti. Les « fantômes » sont de retour, l'« esprit » est revenu.

2

Cependant, reconnaître l'importance des concepts mentaux, accueillir avec bienveillance le retour de l'« esprit » (ou plutôt admettre qu'il n'était jamais véritablement parti) ne signifie nul-

lement qu'il soit indispensable d'adopter les thèses néo-mentalistes, très en vogue actuellement, ni de faire du psychologue l'arbitre des questions de sens.

C'est une chose de dire qu'il faut parler de croyances, de désirs, d'intentions, de conscience, si on veut construire une théorie acceptable du sens, c'en est une autre de dire que ces croyances, désirs, intentions, états de conscience existent réellement, et qu'une étude naturaliste, physique, pourra nous faire progresser dans notre compréhension de ce que c'est que signifier.

Conclure que c'est à la psychologie expérimentale (ou aux neurosciences) que doit être transmise la charge de faire progresser cette théorie, ajouter, comme on le fait habituellement désormais, qu'il est temps de passer du sociologique au psychologique, c'est, à mon sens, faire fausse route, tirer une conclusion qui n'est nullement impliquée par les prémisses.

Il me semble que c'est exactement la conclusion inverse qu'il faudrait tirer. La nécessité dans laquelle nous nous trouvons de revenir aux concepts de croyance, de conscience, d'intention (et aussi de volonté, désir, etc.) pour élaborer de bonnes théories du sens, devrait nous éloigner du psychologique dans la mesure où l'ensemble de ces concepts dits intentionnels sont normatifs (ce que n'importe quelle inspection d'un compte rendu de recherche psychologique, aussi indulgente soit-elle, montrerait aisément). Ils ne décrivent pas ce qu'on observe : ils orientent l'observation, et ils s'organisent en termes de cohérence, de rationalité, de bien et de mal. En un certain sens, ce sont des concepts *moraux*.

Le behaviorisme n'a plus beaucoup de partisans, alors que le mentalisme doit plutôt se protéger contre ses trop nombreux zéloteurs. Mais le mentalisme, c'est-à-dire la doctrine qui, en affirmant la réalité des états mentaux et leurs pouvoirs causaux, accepte dans leurs grandes lignes les idées pré-théoriques que nous nous faisons du rôle des croyances et des désirs dans l'action¹, le mentalisme est loin d'être une solution de rechange enthousiasmante.

1. Sur la discussion behaviorisme-mentalisme la littérature est devenue tellement abondante que je préfère me contenter de citer deux essais clairs et érudits de Pascal Engel : *Introduction à la philosophie de l'esprit*, Paris, La découverte, 1994, et « La philosophie de l'esprit et des

Le mentaliste devrait être en mesure de justifier l'idée que les états mentaux sont des entités réelles, c'est-à-dire, dotées de *pouvoirs causaux* et de *conditions d'identité* bien déterminées. Mais c'est précisément ce qu'il n'a jamais réussi à faire de façon convaincante.

Le problème de la causalité mentale est assez bien connu, car il en existe depuis longtemps de nombreuses versions, sous diverses appellations : problème de l'épiphénoménalisme ou problème du dualisme âme-corps ou des relations corps-esprit, etc.

La question des conditions d'identité des états mentaux est un peu moins discutée. On a essayé d'identifier les états mentaux à partir du langage. Mais on s'est vite trouvé confronté à une difficulté : avons-nous les moyens de montrer que ces états mentaux ont une structure aussi complexe et élaborée que celle du langage ? Un lexique ? Des règles de composition ? Une durée ? Des parties ? Aucune étude empirique sérieuse n'a été capable d'exhiber cette structure : elle est simplement postulée, dans la thèse du « langage de la pensée », sur la base d'une analogie supposée avec l'ordinateur. D'ailleurs, même si cette analogie était fondée, elle mettrait le mentaliste dans une situation embarrassante, car elle ferait du mentalisme une variété du behaviorisme, en allant à l'opposé de ses ambitions initiales.

Une autre façon de donner les conditions d'identité des états mentaux consiste à les individuer à partir de l'environnement physique, en postulant l'existence de relations causales entre les objets et les états mentaux et de variations concomitantes entre les deux. Mais en substituant une relation causale à la relation intentionnelle, on fait des états mentaux des réactions à des stimuli, ce qui revient, encore une fois, à endosser une forme de behaviorisme anti-mentaliste.

Étant donné que les pouvoirs causaux des états mentaux et leurs conditions d'identité sont douteux, leur réalité ne peut pas être rigoureusement affirmée et le mentalisme est menacé.

.... sciences cognitives » dans *La philosophie anglo-saxonne*, Michel Meyer, éd., Paris, PUF, 1994, p. 529-583 ; et les livres d'Élizabeth Pacherie (*op. cit.*, 1993) et de Fabien Cayla, *Routes et déroutes de l'intentionnalité*, *La correspondance R. Chisholm-W. Sellars*, Combas, L'éclat, 1991, qui, défendant des points de vue différents, proposent aussi d'autres références.

Il faut espérer qu'il peut y avoir quelque chose entre le mentalisme et le behaviorisme ou au-delà de ces deux doctrines.

Ce qu'on appelle l'interprétationnisme ressemble à cette troisième voie¹ (compatible avec la plupart des théories sociologiques). Il existe plusieurs variétés d'interprétationnisme.

D'abord les instrumentalistes étroits : ils estiment que l'attribution d'intentions, de croyances, de désirs est une fiction utile parce qu'elle nous permet de rendre intelligibles certains comportements (de l'homme, mais aussi des animaux, des machines, etc.).

Ensuite, ceux qu'on pourrait appeler expressivistes. Ces derniers se gardent bien d'étendre aux animaux et aux machines l'attribution d'intentions, de croyances (et même, en un certain sens, de désirs). Et s'ils semblent manifester une certaine sympathie pour l'idée du gain d'intelligibilité, c'est tout à fait différemment de l'instrumentaliste. Pour l'expressiviste, il est absurde de donner un sens instrumental à l'idée de gain d'intelligibilité. Nous attribuons des pensées, des croyances, des intentions (et même des désirs) aux êtres doués de la capacité de parler, d'utiliser un langage et à eux seulement, parce que nous ne pouvons pas faire autrement, quels que soient les avantages et les inconvénients (on peut donner différentes significations à ce « parce que nous ne pouvons pas faire autrement », conceptuelles, naturelles, etc., mais l'argument demeure). La position expressiviste nous permet d'exclure l'attribution d'intentions ou de croyances aux oranges, aux poumons, aux ordinateurs et même aux vaches et aux poissons, sans tomber dans les difficultés du réalisme des états mentaux. Ce n'est que dans le contexte de nos relations à des êtres parlants que le problème des fictions intentionnelles se poserait.

1. Sur la possibilité d'une « troisième voie » entre le mentalisme et le behaviorisme, il faut lire, bien sûr, tout ce qui a été dit des positions de Wittgenstein à ce sujet, (cf. par exemple Jacques Bouveresse, *Le mythe de l'intériorité. Expérience, signification et langage privé chez Wittgenstein*, Paris, Minuit, 1976 (2^e éd. avec une nouvelle préface, 1987) ; mais aussi ce que proposent des philosophes comme Wilfrid Sellars (cf. F. Cayla, *op. cit.* 1991), Donald Davidson (*op. cit.* 1993), Daniel Dennett, *La stratégie de l'interprète*, (1987), trad. P. Engel, Paris, Gallimard, 1990. Sur ce sujet, les remarques de Vincent Descombes, dans *La denrée mentale*, Paris, Minuit, 1995, sont extrêmement précieuses.

Mais l'instrumentaliste et l'expressiviste ont tout de même quelque chose en commun ; ils semblent penser que l'attribution de désirs, de croyances, d'intentions a deux fonctions principales : prédiction et intelligibilité. Ils sous-estiment une troisième fonction pourtant très évidente : la justification du jugement moral.

Le jugement moral (évaluation et prescription) exige l'attribution à autrui d'intentions, de pensées (pour juger de sa bonne ou mauvaise volonté par exemple, ou pour l'estimer agent ou responsable, ou au contraire victime et irresponsable). Il me semble que cette troisième fonction de l'attribution des croyances et des intentions devrait être réhabilitée. C'est qu'on peut fort bien envisager des prédictions réussies, sans aucune « stratégie intentionnelle » (le behaviorisme le plus plat est à l'origine des prédictions les plus efficaces). Les choses sont un peu plus complexes lorsqu'il est question de l'intelligibilité du langage. On peut estimer que l'attribution d'états mentaux (à soi-même ou à autrui) n'est requise que lorsqu'on délaisse l'étude des phrases types (de la langue, comme on dit) au bénéfice des occurrences, des phrases utilisées en situation de communication (discours, parole, énonciation, etc.). Naturellement, ceux qui résistent à cette dichotomie jugeront que l'analyse du sens exige toujours l'attribution d'états mentaux (réaliste ou fictionnaliste).

Il est un contexte dans lequel ce genre de problèmes se pose de façon tout à fait différente, un contexte dans lequel l'attribution d'états mentaux, d'une « vie intérieure » semble nécessairement requise ; c'est celui du jugement moral (tel que nous le concevons). Autrement dit, ce n'est que dans le contexte moral que l'intentionnalité ne peut pas être éliminée. C'est dans ce contexte que devraient être puisés les meilleurs arguments en faveur de l'interprétationnisme.

Même si on pouvait trouver les moyens de prédire mieux sans le concours des concepts intentionnels, cela ne veut pas dire que nous serions tentés de les abandonner. Car abandonner ces concepts reviendrait à ruiner la possibilité du jugement moral.

Il ne semble pas tout à fait exclu qu'on puisse arracher la « vie intérieure » aux impasses du mentalisme étroit, sans la dissoudre complètement dans une forme de behaviorisme aujourd'hui discredité.

En tout cas, c'est ainsi qu'il faut procéder, me semble-t-il, dans l'analyse des sentiments où le mentalisme est une tentation constante et le behaviorisme une menace persistante.

3

La « redécouverte » des sentiments en général et celle de sentiments particuliers tels que la haine, la honte, la peur, etc. semble participer de ce mouvement général de réhabilitation de concepts dont les behavioristes pensaient qu'il valait mieux les ranger dans le musée du folklore psychologique¹. Mais l'analyse des sentiments est, pour des raisons assez évidentes, plus atteinte encore par la dérive mentaliste.

La première tentation, bien sûr, c'est de réduire les sentiments à des états purement qualitatifs, aux impressions que le sujet « ressent ». Mais on voit très vite que le critère purement qualitatif (ce que le sujet ressent) n'est pas suffisamment fin. Il est assez évident que ce qui nous permet de distinguer l'envie, la jalousie, la haine, la rancune, le ressentiment, la répulsion, ce n'est certainement pas un certain type d'impression ressentie par un sujet.

On passe alors à des critères plus cognitifs. On range les sentiments du côté de la pensée, plutôt que des sensations. Le sentiment devient une attitude intentionnelle dirigée vers un objet (propositionnel ou nominal: pensée, croyance, personne). Il n'est plus une simple réaction, une passion causée par des objets. En réalité, on n'a pas beaucoup avancé. Le passage de la passion à l'action, de l'affect à l'intention, de la sensation à la pensée ne permet pas de répondre à la question de l'identification ou de la discrimination des sentiments. Ni l'objet propositionnel de l'attitude émotive (ce qu'on hait, ce qu'on redoute, etc.), ni le style de comportement ne permettent de comprendre comment nous réussissons, dans la vie quotidienne, à identifier ces gammes subtiles des sentiments.

En définitive, il semble qu'il soit nécessaire de faire intervenir l'idée d'une procédure sociale d'attribution des sentiments pour

1. Rorty A.O., éd., *Explaining Emotions*, Berkeley, University of California Press, 1980.

comprendre comment nous faisons ces fines distinctions, et tout aussi nécessaire d'introduire l'idée d'une dimension morale dans cette procédure.

L'idée que les sentiments ne peuvent être reconnus ou identifiés qu'en tant qu'objets moraux ou sociaux a d'innombrables conséquences épistémologiques. Si les sentiments ne sont pas à proprement parler des entités « réelles » ou « naturelles » mais des objets socialement construits, cela signifie, entre autres, qu'on ne peut pas leur laisser un rôle causal ordinaire, ce rôle qu'ils semblent avoir dans certaines explications extraordinairement simplistes de phénomènes aussi dramatiques que la xénophobie, les conflits ethniques, etc. (il suffit de songer au rôle que joue la « haine » dans ces explications, sans qu'une définition de ce qu'elle est soit donnée, même vaguement).

La question de savoir si les sentiments peuvent être soustraits à l'approche étroitement mentaliste est donc à l'ordre du jour.

2. LA HONTE

La honte est-elle un sentiment primitif, indéfinissable qu'on ne peut pas connaître si on ne l'a pas éprouvé, de la même façon qu'on ne peut pas connaître le goût des groseilles si on n'en a pas mangées ?

À première vue, oui. Ou plus exactement, il semble que ce que nous disons de nos hontes présente les traits caractéristiques des énoncés « à la première personne » :

1) l'incorrigibilité : je ne peux pas répondre raisonnablement à celui qui me dit « J'ai honte », « C'est faux, tu n'as pas honte ».

2) la certitude : je ne peux pas dire raisonnablement « Je crois que j'ai honte, mais je n'en suis pas sûr. Je vais vérifier » (de la même façon que je ne peux pas dire « Je crois que j'ai peur, mais je n'en suis pas sûr, je vais vérifier »).

Même s'il n'y a pas beaucoup de sens à parler de description du sentiment de honte, dès lors que la possibilité de l'erreur est exclue (une objection de type wittgensteinien), on peut néanmoins lui donner une expression au moyen de métaphores. C'est ce qu'on fait habituellement lorsqu'on présente la honte comme une sorte de maladie dont on pourrait évaluer la progression : de la honte légère (on se « sent » simplement gêné, embarrassé, noué) à la honte aiguë (on se « sent » étouffé, anéanti, pulvérisé) — les nosographies pouvant être plus ou moins subtiles et les nuances plus ou moins variées.

En contradiction avec ce point de vue interne, subjectif, à la « première personne », on en a proposé d'autres plus ou moins externes, objectifs, à la « troisième personne ». On a d'abord tenté d'associer ces sensations internes relativement indéterminées à des manifestations externes ou comportementales caractéristiques : rougeur plus ou moins étendue ou accentuée, et, dans les cas extrêmes, pâleur et grelottement ; posture de la tête baissée ou détournée ; dissimulation de la face par les mains ; mimiques caractéristiques d'hébétude ou de désolation ; tendance à fuir la situation, à éviter les témoins ; réactions de colère ou de rage après coup et ainsi de suite.

Différentes théories behavioristes se contentent de réduire ces sentiments à ces manifestations externes, en laissant ouverte la possibilité qu'elles soient accompagnées ou non de sensations internes non caractéristiques. Elles ne se prononcent pas sur la forme que peuvent prendre ces événements mentaux internes, qui échappent, par définition, à la mesure, à l'observation.

D'autres théories rejettent ce genre de behaviorisme étroit, en établissant une sorte de lien logique entre les manifestations externes et internes ou en niant la valeur de la dichotomie interne-externe¹.

Mais il n'est pas nécessaire d'être un behavioriste étroit ou un adversaire déterminé de toute dichotomie entre l'interne et l'externe pour contester l'idée que la honte soit purement et simplement un sentiment indéfinissable, auquel il existerait un accès privilégié, à la « première personne ».

Ce qui nous interdit de réduire la honte à ce que nous « ressentons » tient plutôt à la disparité entre d'une part, la richesse de notre vocabulaire sentimental, la finesse de nos discriminations conceptuelles en matière de sentiments et, d'autre part, la pauvreté de ce que nous disons ou pouvons dire de nos états qualitatifs, de nos impressions.

Il n'est pas du tout sûr que nous aurions les moyens de distinguer la honte de la peur, de la tristesse, de la colère ou de la rage par exemple, si nous étions forcés de nous en tenir à ce que nous « ressentons ».

Je ne vais pas essayer de le prouver, à proprement parler, mais seulement de montrer la complexité de ce que nous savons dire de la honte, une complexité telle qu'elle semble nous interdire toute réduction de la honte au sentiment.

De ce point de vue, l'opposition entre avoir honte *de* quelque chose ou *de* quelqu'un et avoir honte *pour* quelqu'un est certainement cruciale. Je ne vois pas très bien comment on pourrait distinguer ces deux hontes si on s'en tenait à ce qu'on ressent (la même chose désagréable dans les deux cas probablement). En défendant cet argument, je ne m'éloignerai pas trop de Hume, en dépit de ce qu'on dit souvent.

1. Tous ces sujets sont abordés dans le recueil d'Amélie O. Rorty, *op. cit.*, 1980.

De nombreux critiques et exégètes estiment que Hume est « subjectiviste » et que, dans sa théorie des passions, il accorde la plus grande place aux impressions, à ce qu'on ressent¹.

Cependant, ce qui, d'après Hume, nous permet de distinguer la haine de la honte ou l'amour de l'orgueil, ce n'est certainement pas une impression ressentie, un état qualitatif ou « vécu ». Ce qu'on ressent dans la haine et la honte est approximativement identique (un vague sentiment de déplaisir) et ce qu'on ressent dans l'amour et l'orgueil est aussi à peu près identique (un vague sentiment de satisfaction). Mais ce qui distingue la haine de la honte, c'est le fait que la haine est dirigée vers autrui et la honte vers soi, et ce qui différencie l'amour de l'orgueil, c'est que l'amour est dirigé vers autrui et l'orgueil vers soi. Ce qu'on ressent n'est donc pas un critère de discrimination suffisant et, même s'il est nécessaire, il reste vague.

C'est aussi évident lorsque les oppositions se font à l'intérieur du domaine de la honte et c'est pourquoi j'accorde tant d'importance à l'opposition honte *de* — honte *pour*.

Mais avant de discuter cette question, je vais présenter d'autres divisions conceptuelles qui me font penser que la honte ne peut décidément pas être comprise simplement du point de vue de la première personne. Et je conclurai sur ce qui, plus que tout autre argument, plaide en faveur d'une analyse non étroitement affective de la honte : le fait qu'elle a des raisons et non des causes.

1

À la lecture de nos bons dictionnaires, il apparaît qu'il existe, en gros, deux usages du substantif « honte ».

La honte est « un effet d'opprobre entraîné par un fait, une action transgressant une norme éthique ou une convenance (d'un groupe social, d'une société) ou par une action jugée avilissante par rapport à une norme (d'un groupe social, d'une société) ». Mais c'est également un « sentiment de pénible humiliation qu'on

1. Davidson a mis en évidence ce genre de préjugé philosophique à l'égard de Hume et il s'est efforcé de montrer leur inanité dans « La théorie cognitive de l'orgueil de Hume », (1976) dans *op. cit.*, 1993, p. 365-383.

éprouve en prenant conscience de son infériorité, de son imperfection (vis-à-vis de quelqu'un ou de quelque chose) » (*Trésor de la Langue française*).

Ce que le dictionnaire veut dire, c'est que « honte » indique ou bien un sentiment (ce qui se passe en moi au moment où j'ai honte) ou bien un blâme (auquel correspond un sentiment qui n'est pas la honte, mais l'indignation).

« J'ai honte » décrit la honte. « C'est une honte » rapporte directement un acte, un jugement, un blâme et décrit indirectement l'indignation.

Mais cette séparation est trompeuse. En effet, les deux hontes s'impliquent. Il est impossible de définir l'une sans avoir l'idée de l'autre. Toutes les transgressions aux convenances ne sont pas sanctionnées par le blâme « c'est une honte ». La violation d'une règle peut être « ignoble », « ignominieuse », « odieuse », « scandaleuse », « grotesque », « ridicule », « gratuite », « mesquine », « indécente » ou « honteuse ». Il existe un barème de peines, une hiérarchie de ces sanctions informelles. « C'est une honte » semble situé entre « l'ignoble » et le « ridicule ». Mais ce barème n'est pas clairement gradué.

En réalité, le seul moyen de savoir si le blâme « c'est une honte » est un blâme approprié, c'est de supposer que la peine frappant celui qui a enfreint la règle devrait être la honte (pas plus que la honte et pas moins). La spécificité du blâme « c'est une honte » (par opposition à « c'est ignominieux » ou « c'est mesquin ») disparaîtrait, si nous ne prenions pas comme mesure le sentiment de honte que la personne blâmée devrait éprouver. Nous avons donc ce cercle (que le dictionnaire masque à peine). Qu'est-ce qui est « honteux » ? C'est ce qui devrait provoquer la honte. Qu'est-ce la honte ? C'est le sentiment qu'on *devrait* éprouver lorsqu'on accomplit une action « honteuse ».

Nos dictionnaires estiment qu'on peut dissiper la confusion en opposant un usage personnel (« j'ai honte ») et un usage impersonnel (« c'est une honte »). Dans le premier cas, on évoque le sentiment ; dans le second, on évoque le blâme ou la sanction. Mais on ne peut pas s'en sortir aussi facilement, hélas.

Lorsque je dis à mon créancier « J'ai honte de ne pas pouvoir te rembourser », ce n'est pas seulement pour décrire ou évoquer mon sentiment. Mon créancier n'est pas un psychologue impar-

tial qui s'intéresse à ma vie intérieure. En disant « J'ai honte de ne pas pouvoir te rembourser », j'exprime un jugement négatif sur le fait que je n'ai pas remboursé mes dettes, un jugement qui devrait en principe servir d'excuse (car il manifeste indirectement une bonne intention) ou, au moins, de réparation provisoire à l'offense. Bref, en disant « J'ai honte de ne pas pouvoir te rembourser », je ne me contente pas de décrire mon sentiment interne, je qualifie de façon négative un certain comportement (le non-remboursement de la dette). C'est une forme de sanction.

Les raisons ne manquent pas, qui devraient donner à la honte-sanction la priorité sur la honte-sentiment. L'idée qu'il existe quelque chose comme un sentiment de honte spécifique et clairement distinct de la colère ou de la peur n'est peut-être qu'une idée dérivée de notre conception de ce qui est honteux.

2

Dans certains cas, il n'est pas très facile de faire correspondre à l'expression « J'ai honte », un sentiment spécifique et cela ne tient pas seulement à la maîtrise insuffisante du concept de « honte ».

Lorsqu'on dit « J'ai honte de montrer mes notes lamentables à mes parents », il y a, semble-t-il, confusion ou recouvrement entre la honte et d'autres sentiments apparentés tels que la crainte ou la peur. D'ailleurs les Anciens étaient sensibles à cette confusion, puisqu'ils recommandaient de distinguer deux formes de honte.

Habituellement, nous imaginons la honte comme un éclair de conscience, une illumination soudaine, un retournement brutal qui transfigure ou plus exactement enlaidit très fortement notre situation, en jetant une lumière glauque sur nos actes ou sur la considération dont nous pensions être l'objet. La honte est une forme de déception¹. Ce n'est pas du tout ainsi que les Anciens s'imaginaient la honte. Comme ils étaient sans doute beaucoup plus vertueux et beaucoup plus énergiques que nous, ils concevaient la honte comme une « émotion tranquille » réglant le cours

1. Guillaumin J., « Honte, culpabilité et dépression », *Revue française de psychanalyse*, 5-6, XXXVII, 1973.

de nos actions, retenant les mouvements blâmables, les étouffant avant qu'ils adviennent. La honte pour eux, ressemblait à ce que nous appelons « pudeur ». Pour les Anciens, la honte était accessoirement la punition du vice, mais c'était principalement la meilleure des vertus. La vraie honte, c'était la bonne honte, cette crainte qui nous fait fuir le mal. La honte était une « crainte divine » : une crainte qu'on éprouve à l'idée de ne pas plaire à nos amants, une crainte qu'on éprouve à l'idée de faire ce qu'il ne faut pas faire¹.

3

Pour distinguer la honte de la culpabilité, on dispose de quelques faits grammaticaux, si on peut dire. Une bonne explication de ces faits passe, me semble-t-il, par la compréhension du rôle de la honte-sanction. Voici les faits.

On ne dit pas « J'ai honte mais je suis coupable » (« J'ai honte de l'état de ce phoque ; mais c'est moi qui l'ai mis dans cet état »). Mais on dit « Je suis coupable mais j'ai honte » (« Je suis coupable d'avoir torturé un phoque, mais j'ai honte de ce que j'ai fait »). Et voici une interprétation en termes de honte-sanction.

La honte s'oppose à la culpabilité ; elle l'atténue. C'est un début de réparation. On peut donc dire : « Je suis coupable, mais j'ai honte ». Dire : « Je suis coupable mais j'ai honte » n'est pas moins acceptable que dire : « Je suis avare, mais je donne aux pauvres ».

Si la conjonction « mais » nous semble appropriée, c'est qu'il y a une opposition ou une inversion dans l'argumentation. Ce qu'on pense de la honte s'oppose à ce qu'on pense de la culpabilité (de la même manière que ce qu'on pense du fait de donner aux pauvres s'oppose à ce qu'on pense de l'avarice).

Mais ce qu'on pense de la culpabilité ne s'oppose pas à ce qu'on pense de la honte. Dire : « J'ai honte mais je suis coupable » est aussi absurde que dire : « Je suis avare mais je garde tout mon argent pour moi dans une cassette bien cachée ». La culpabilité s'ajoute à la honte, comme un accent : « J'ai honte et en plus je suis coupable ».

1. Platon, *Lois*, I, 646, E, ed. et trad. E. Des Cases, Paris, Les Belles Lettres, 1951.

La honte est une réparation pour ce dont nous sommes coupables. Mais la culpabilité n'est pas une réparation pour ce dont nous avons honte.

4

On peut dire : « J'ai honte *de* mon nez », mais on ne peut pas dire : « J'ai honte *pour* mon nez ». Pourquoi ?

On peut avoir honte *de* n'importe quoi : un objet (une chemise tachée, un appartement en désordre), une partie de personne (un nez, un genou) une action (une gaffe) etc.

On ne peut avoir honte *pour* qu'à l'égard d'une personne. On ne peut pas avoir honte *pour* sa chemise, *pour* son genou, *pour* sa gaffe. Mais, c'est parce que seules des personnes peuvent avoir honte, jamais des objets ou des parties de personne. Je peux avoir honte *de* mon appartement en désordre, mais mon appartement n'aura jamais honte, même s'il est en désordre. Je peux avoir honte *de* mon genou, mais mon genou n'aura jamais honte, aussi affreux soit-il.

De ceci il semble suivre que la honte *pour* quelqu'un est beaucoup plus proche du blâme que la honte *de* quelqu'un ou *de* quelque chose. En effet, blâmer quelqu'un par la honte, c'est suggérer qu'il devrait avoir honte. On ne peut jamais suggérer à un nez ou à un appartement d'avoir honte, parce qu'il n'est pas possible que des parties de personnes ou des objets aient honte.

Mais on pourrait se contenter de dire que la honte *pour* ressemble à un blâme adressé à autrui, alors que la honte *de* est plus proche d'un blâme adressé à soi-même.

Lorsqu'on a honte *de*, ce dont on a honte appartient au sujet de la honte (J'ai honte *de* : mon nom, mon nez, mes oreilles, ma cuisine, mes fautes d'écriture, mes sottises. Mais j'ai également honte *de* : mes parents, mes amis, mes alliés, etc.). Les choses et les personnes deviennent équivalentes. Ce sont des possessions (qui me déprécient).

Lorsqu'on a honte *pour*, ce pourquoi on a honte n'appartient pas au sujet de la honte. Les clowns qui sont incapables de faire rire, les chanteurs qui n'arrivent pas à chanter juste, les acteurs qui jouent faux, les danseurs qui gesticulent à contretemps me

font tantôt honte, tantôt pitié. Lorsque j'ai honte, j'ai honte *pour* eux, sans les considérer comme des possessions qui me déprécient (j'en ai déjà suffisamment).

Dans la honte *de*, on ne blâme jamais un autre que soi-même. Dans la honte *pour*, on est censé éprouver une peine pour un autre que soi. Cette peine, on peut s'en défaire pour la laisser à l'autre. La honte *pour* quelqu'un est toujours au bord du sarcasme ou au bord de la pitié.

Ce qui distingue la honte *de* et la honte *pour*, ce n'est pas la qualité de l'affect qui est ressenti (rien ne nous interdit de penser que c'est exactement la même chose qu'on ressent) ; c'est plutôt leur inscription contextuelle, leur destin social, ainsi qu'on peut s'en convaincre en comparant les exemples suivants :

1) Le petit Michel Leiris étudie au lycée Jeanson de Sailly, dont les élèves appartiennent en majorité à des familles beaucoup plus riches que la sienne. Un soir, sa mère vient le chercher à la sortie du lycée. Le lendemain, un de ses condisciples lui demande si la personne qui est venue le chercher la veille est sa « mère ou son institutrice ». « C'est mon institutrice » répond le petit Michel Leiris, qui, devenu grand, conclut : « Grande aujourd'hui est ma honte d'avoir ainsi trahi, bouche tordue par une honte celle-là des plus mesquines, l'être à qui le garçon à peine sorti de sa coquille que j'étais en ce temps aurait dû se sentir lié plus qu'à tout autre sans avoir eu à formuler le moindre serment¹. »

2) Un modeste écrivain raconte son récent voyage dans un pays en crise. Devant les journalistes qui l'interrogent, il s'égare dans une digression à propos des « queues » de consommateurs. Les consommateurs sont vite oubliés, et l'écrivain parle des « queues » qu'il classe subtilement selon leur longueur, leur épaisseur, leur flexibilité, leur sensibilité. Les journalistes, embarrassés, essaient d'interrompre l'écrivain visiblement inconscient du caractère tendancieux de son récit. Mais la verve classificatrice de l'écrivain l'emporte, et les journalistes impuissants (si on peut dire) doivent renoncer à le couper (si on peut dire). Les journalistes rougissent un peu, baissent la tête, s'agitent légèrement, se regardent ; ils se couvrent les yeux et finissent par sourire, alors que le suicide moral de l'écrivain se déroule, interminable, devant eux.

1. Leiris M., *À cor et à cri*, Paris, Gallimard, 1988, p. 66.

Renoncez un peu à la psychologie et vous verrez que ces deux scènes se distinguent surtout par leurs aspects « tactiques ». Les journalistes peuvent faire alliance contre l'écrivain ridicule alors que le petit Leiris ne peut nouer aucune alliance, ni du côté du camarade, ni du côté de la mère.

Observez les journalistes : ils baissent la tête, puis se regardent. Ils rient sous cape. Ils se sont faits complices contre la victime. Ils ont noué une alliance. Leur honte s'est évanouie. Mais celui qui a honte *de* quelqu'un est seul. Le honteux *de* est trop solidaire de la personne dont il a honte et pas assez étranger au témoin pour se moquer de son regard. Le honteux *pour* n'est pas nécessairement solidaire de la personne dont il a honte : il peut être suffisamment complice avec les autres témoins pour fabriquer une petite communauté de médisants. La honte *de* et la honte *pour* n'ont donc pas le même destin. Elles se distinguent par leurs issues tactiques, par les alliances qu'elles autorisent et celles qu'elles interdisent.

D'autre part, la honte *pour* est plus évidemment une honte-sanction. La honte *pour* quelqu'un se comprend habituellement :

1) psychologiquement (c'est un processus d'identification « empathique » : On éprouve ce que l'autre éprouve. On se met à la place de l'autre) ;

2) sociologiquement (il y a une certaine quantité de honte à offrir en réparation à l'offense contre le bon goût ou la tranquillité de l'audience).

À première vue, la seconde explication est plus convaincante, car lorsqu'on a honte pour quelqu'un, c'est la plupart du temps parce que quelqu'un n'éprouve pas la honte qu'il *devrait* éprouver. Apparemment, celui pour qui j'ai honte est trop préoccupé à se couvrir de ridicule pour avoir le temps d'éprouver de la honte. Je prends donc sur moi cette honte qu'il *devrait* éprouver. Je me dévoue, en quelque sorte.

Il faut verser le prix de la honte et l'identité du payeur n'a pas d'importance. La honte nous prend tous en otage. Je me sacrifie pour sauver la scène. Pour avoir une idée moins pathétique de ce que représente ce sacrifice, imaginons un convive relançant la conversation lorsqu'un « ange passe ». Il y a un long silence autour de la table ; tout le monde est gêné. Quelqu'un doit se sacrifier au nom des intérêts supérieurs de la conversation. Il

importe peu que ce soit ce convive-ci ou cet autre (c'est, hélas, souvent le même qui « sauve » la situation). Prendre l'initiative de relancer la conversation c'est montrer sa solidarité à l'égard des convives embarrassés. En dépit de ces sacrifices, les conversations sauvées ne sont pas toujours de bonnes conversations. Si un animateur passe sa soirée à « boucher les trous », si on peut dire, c'est que quelque chose ne va pas. On imagine l'animateur comme une sorte de saint, toujours prêt à se mettre à l'ouvrage pour accomplir le bien. C'est trop lui accorder. Car, souvent, c'est l'animateur impatient qui détruit la soirée : il accapare la conversation ; il relance à tout propos ; il brise la quiétude des convives. En réalité, il invente des problèmes qui n'existent pas. Ce qui signifie qu'il n'est pas un saint, mais un pécheur. Car c'est lui qui jette la pierre. Par ses mouvements intempestifs, l'animateur impose une image très triste, très négative de la conversation. Par ses relances répétées, l'animateur semble dire que les autres sont, d'une certaine façon, trop médiocres ou trop endormis pour se débrouiller sans lui. Son sacrifice est une manière de blâme ou une forme d'orgueil. (« Ça ne va pas ; on s'ennuie, voilà ce que vous devriez faire. Ah, si je n'étais pas là ! ») La même ambivalence existe lorsqu'on a honte *pour* quelqu'un. En me montrant honteux, je montre à l'autre la peine qu'il *devrait* ressentir. Ce n'est pas très aimable de ma part. Si je conçois ma honte *pour* l'autre comme un sacrifice, c'est, de toute façon, un sacrifice didactique, un sacrifice plein de ressentiment. (« Regarde, ce que je dois faire à cause de toi. Regarde ce que tu devrais faire si tu n'étais pas aveugle ».)

Parler de sacrifice, dans ces conditions, est donc un peu exagéré. Rien ne nous interdit, au demeurant, d'envisager la possibilité que celui qui a honte *pour* quelqu'un voit des ridicules qu'il est le seul à remarquer. En ce sens, c'est lui qui commet une faute en étant trop attentif aux fautes de l'autre. Bref, avoir honte *pour* quelqu'un, ce ne serait vraiment rien d'autre qu'une façon de le blâmer, même si ce blâme est « intériorisé »¹, pour

1. C'est le genre d'hypothèse qu'on pourrait formuler à partir de la théorie des émotions de Jennifer Church : « L'émotion et l'intériorisation des actions », trad. Yves Bellanger dans *Raisons Pratiques*, 6, (*La couleur des pensées*), Paris, E.H.E.S.S., 1995.

ainsi dire, non exprimé publiquement dans un sarcasme cruel ou un grand éclat de rire. Avoir honte *pour* quelqu'un, ce serait comme un rire étouffé.

5

Si nous pensons que la générosité étalée n'est plus la générosité, ce n'est pas en vertu d'une loi qui dirait : « La générosité diminue d'autant plus qu'on l'étale ». On ne découvre pas, stupéfait, que le sentiment généreux s'évanouit à mesure qu'on l'affiche. La disparition de la générosité par l'étalage n'est pas un fait qu'on apprend par introspection ou par observation. C'est simplement une disparition logique, si on peut dire, par contradiction. L'idée de générosité est incompatible avec celle d'ostentation. Rien ne peut être à la fois généreux et ostentatoire, car la générosité est offrande et don sans retour, alors que l'étalage est satisfaction d'amour-propre, c'est-à-dire salaire, gain, bénéfice.

En va-t-il de même pour la honte ? La disparition de la honte par son exhibition volontaire est-elle l'effet d'une sorte de contradiction dans les termes ?

À première vue, ce qu'on appelle « sentiment » ne se détruit pas à la manière d'une vertu, par une sorte de contradiction interne, une faute logique.

Soit la joie : vous croyez avoir tiré le bon numéro au loto, ce qui, bien entendu, vous fait exploser de joie. En consultant votre billet, vous vous apercevez que cette croyance était fausse : votre numéro n'était pas le bon, à un chiffre près. Votre joie explosive va probablement s'évanouir. Car un sentiment disparaît, semble-t-il, avec la disparition de la croyance qui en est la « cause ».

Ce n'est pas une relation interne, une « faute logique » qui détruit le sentiment, mais une relation externe (la disparition de la croyance que j'ai gagné au loto dans ce cas précis). C'est du moins ce qu'on a tendance à penser. Mais il est clair, aussi, qu'un sentiment peut disparaître, lorsque la signification de la cause est subitement transformée.

Soit une action malveillante qui « cause » votre colère ou votre ressentiment. On vient de vous pincer dans un ascenseur. Vous éprouvez une juste colère. Elle peut s'atténuer ou disparaître si l'offenseur (le pinceur) s'excuse. On l'a bousculé, il a trébuché,

etc., il s'est rattrapé en vous pinçant. C'est difficile à croire, mais si vous préférez ne pas vous disputer, vous pouvez accepter cette excuse. L'offenseur s'est contenté de transfigurer les causes. Ce qui, pour vous, était une action volontaire, une attaque délibérée est devenu, par la grâce des excuses, un geste dépourvu de toute intention malveillante, comme un éternuement. On a pincé ; ça a pincé ; c'est comme « il pleut », il n'y a plus de coupable. La conversion d'une action en événement, dégonfle, si on peut dire, votre colère. Celle-ci était provoquée, causée par une action « intentionnelle » ; l'action « intentionnelle » a cessé de l'être par les excuses ; la « cause » de votre colère a disparu. Inversement, la conversion d'un événement en action peut provoquer le même miracle.

Soit la honte : vous venez de casser un vase de Chine par un geste maladroit. Revendiquez le geste ; dites : « J'ai l'habitude de casser les vases de Chine », personne ne vous croira, mais il ne sera plus question de votre honte ; plutôt de votre stupidité ou de votre arrogance.

En transfigurant un mouvement involontaire en geste délibéré, en passant de l'événement à l'action, vous supprimez la honte, pour ainsi dire. C'est un genre de miracle qu'on observe rarement dans la nature où l'événement reste ce qu'il est, indépendamment de ses descriptions et où ses effets sont identiques quelles que soient nos interprétations. Est-il donc si raisonnable de parler des causes de la honte purement et simplement ?

L'idée que la honte serait une sorte de réaction à des causes extérieures est assez populaire. On dit qu'on a honte parce qu'on se croit trop petit ou trop grand, parce qu'on est publiquement humilié, parce qu'on est d'origine trop haute ou trop basse, parce qu'on a des parents stupides, parce qu'on a raté un examen facile, parce qu'on a renversé la sauce sur le costume blanc du voisin, etc.

Mais on a du mal à savoir si ces énoncés causaux ne recouvrent pas en réalité des croyances « dispositionnelles ». Tout le monde serait prêt à admettre que la même « cause » extérieure n'a pas le même effet si le sujet qui en subi l'effet n'est pas prédisposé à le subir.

Les savants nous recommandent en effet un peu de calme. Le foisonnement des causes n'est qu'une illusion. En réalité, toutes

ces causes prochaines peuvent être réduites à une seule cause simple et lointaine. Le sentiment d'infériorité, ou comme on le disait naguère, le tempérament honteux. Les propriétés de ce tempérament sont telles qu'on trouvera d'autant plus de raisons d'avoir honte que notre tempérament sera plus honteux. Autrement dit, si on a un fort tempérament honteux, un rien suffira à déclencher la honte (un rien peut signifier : quelque chose d'inexistant).

C'est une façon de réduire considérablement le poids des « causes » extérieures dans le déclenchement de la honte. Parler d'une honte provoquée par telle ou telle cause serait aussi absurde que dire d'un jaloux que sa jalousie est causée par un fait précis quelconque, car si le jaloux finit par être convaincu de l'inanité de ce fait, rien n'exclut qu'il en trouve d'autres, aussi peu convaincants pour les autres et aussi déterminants pour lui. La disproportion est telle entre ces faits et le sentiment qu'elles conditionnent qu'on ne peut pas parler de causes. On peut, à la rigueur, parler d'occasions. Mais c'est déjà trop accorder au monde extérieur. Ce n'est pas parce que les causes sont nombreuses que nous sommes très jaloux : c'est parce que nous sommes jaloux que nous trouvons partout des causes. Rien ne nous interdit de penser qu'il en va de même pour nos hontes. Elles ne sont pas causées par des propriétés objectives de certains événements : elles sont projetées sur ces événements.

D'autre part, on confond souvent la *description d'une relation causale* entre un affect et ce qui le provoque et la *définition causale* d'un affect par ce qui est censé le provoquer. Pour proposer une description convenable d'une relation causale entre un affect et sa cause, il faudrait avoir les moyens de les identifier de façon indépendante ou non circulaire. Or, il arrive souvent que nous n'ayons pas d'autre moyen d'identifier l'affect qu'en décrivant ses causes supposées. Pensez au rire. Lorsqu'on essaie de le définir indépendamment de ses « causes », le résultat est catastrophique. Pour le Robert c'est une expression de gaieté « caractérisée par l'élargissement de l'ouverture de la bouche accompagnée d'expirations saccadées plus ou moins bruyantes » ; et « gaieté » est défini : état ou disposition des êtres qu'anime le plaisir de vivre ou une humeur riante ! Je ne crois

pas que cela suffirait à faire comprendre ce que c'est que « rire » à un Martien dépourvu de cette faculté. Et c'est probablement pour cette raison que les philosophes ont choisi (plus ou moins consciemment) de définir le rire par ses causes sans se soucier du risque de circularité qui ne manquerait pas de les menacer au cas où ils souhaiteraient examiner ensuite les « causes » du rire (qu'est-ce qui fait rire ?).

Selon Hobbes : « La passion du rire est un mouvement subit de vanité provoqué par une conception soudaine de quelque avantage personnel, comparé à une faiblesse que nous remarquons actuellement chez les autres¹. » Le rire est une joie, mais une joie particulière. Une joie causée par la vue imprévue de notre supériorité sur autrui. Il y a beaucoup d'autres joies, qui ne sont pas causées par « la conception soudaine de quelque avantage personnel » et c'est heureux. Mais ces joies ne se manifestent pas dans le rire. Dans la classe des réactions joyeuses, ce qui spécifie le rire, c'est ce qui le « cause ».

Bergson ne s'y prend pas autrement pour « définir » le rire². Quelle est la signification du rire ? C'est ce qui fait rire. Qu'est-ce qui fait rire ? Le mécanique plaqué sur le vivant. Hobbes et Bergson découvrent chacun une cause différente au rire : ce qui veut dire en réalité qu'ils donnent, chacun, une définition différente du rire. Trouvez d'autres « causes », vous trouverez d'autres définitions.

Bref, le langage causal que nous utilisons assez massivement pour parler des affects est plutôt trompeur. Il ne nous permet pas de distinguer clairement les causes et les occasions, les relations causales et les définitions ou les identifications en termes causaux.

C'est pourquoi, d'ailleurs, certains philosophes préfèrent abandonner le vocabulaire des causes et préfèrent s'en tenir au langage des raisons, c'est-à-dire au langage intentionnel (ainsi que nous y invite le fait qu'on a toujours honte de quelque chose ou quelque'un ou pour quelque'un).

1. Hobbes T., *De la nature humaine*, (1651), trad. d'Holbach, Paris, Vrin, 1951, p. 97-98.

2. Bergson H., *Le rire*, (1924) Paris, P.U.F., Quadrige, 1924.

À partir du moment où nous avons choisi d'envisager les prétendues causes de la honte comme des raisons d'avoir honte, nous pouvons comprendre plus aisément comment la honte peut se transfigurer ou disparaître par l'effet d'un processus qui n'a rien de causal et qui ressemble plutôt aux disparitions logiques que nous évoquions à propos de la générosité. Il nous faut seulement établir quelques limites sociales ou logiques à la honte. Commençons par les limites sociales.

Selon Max Scheler, ce qui, pour le missionnaire, est une excellente raison d'avoir honte, n'est pas du tout une bonne raison pour l'« indigène ». L'histoire est pleine de ces pénibles anecdotes. « L'indigène » ne dissimule pas son sexe, mais invitée par le missionnaire « à couvrir des parties honteuses, elle se refuse vivement à le faire, en donnant tous les signes de l'expression naturelle de la pudeur, et dans le cas où elle obéit à contrecœur, elle s'enfuit en courant et se cache derrière des buissons ou dans sa hutte et d'abord on ne peut l'amener en public avec ce cache-sexe¹ ». Pour « l'indigène », porter un short ridicule de missionnaire est une « meilleure raison » d'avoir honte que se promener tranquillement la peau au soleil, c'est bien connu.

Stendhal n'est pas moins relativiste que Scheler sur ces sujets. « Une femme de Madagascar laisse voir sans y songer ce qu'on cache le plus ici, mais mourrait de honte plutôt que de montrer son bras. Il est clair que les trois quarts de la pudeur sont une chose apprise². »

Les limites « logiques » de la honte sont bien plus intéressantes. On ne peut, semble-t-il, avoir honte que pour des actions révolues ou des états de choses présents (des fautes ou des défauts) même si ces actions et ces états de choses n'ont pas eu lieu ou n'existent pas. Mais on ne peut pas avoir honte de l'avenir. On peut craindre l'avenir. On peut craindre d'avoir honte à l'avenir. On peut avoir honte de ce qu'on prépare pour l'avenir. Mais on ne peut pas avoir honte de l'avenir. À moins de le connaître, ce qui le transforme en état de choses présent ou en actions révolues.

1. Scheler M., *La pudeur* (1913), trad. M. Dupuy, Paris, Aubier, 1952, p. 25-26.

2. Stendhal, *De l'amour*, Paris, Garnier-Flammarion, p. 86.

On peut avoir honte *de* presque tout, mais on ne peut pas avoir honte *pour* tout. Je peux avoir honte *de* mes oreilles ; je ne peux pas avoir honte *pour* elles. On ne peut avoir honte que *pour* une personne, jamais *pour* un objet ou une partie de personne. On ne peut avoir honte que *pour* ce qui pourrait ou devrait avoir honte. Ce n'est pas le cas de mes oreilles. On ne peut pas avoir honte de ce qu'on juge bon ou honorable. On ne peut pas avoir honte de ce qu'on revendique. On ne peut pas dire : « J'ai honte d'avoir trahi lâchement mon meilleur ami, mais j'en suis fier ». Il faut choisir. Si on pouvait revendiquer ce qui fait honte, il n'y aurait plus de différence entre la honte et la fierté.

Il serait absurde d'envisager la possibilité qu'il existe, pour chacune des formes conceptuelles de la honte, un état affectif correspondant, une *nuance* de honte bien déterminée. Et il serait tout aussi absurde de vouloir défendre l'idée que ces différentes limites de la honte ressemblent, de quelque façon que ce soit, à des limites physiques. Dire : « Je ne peux pas avoir honte de l'avenir ou de ce que je trouve bon et honorable », ce n'est pas du tout la même chose que dire : « Je ne peux pas voyager dans le temps » ou « Je ne peux pas consommer de l'arsenic à fortes doses et me porter bien ».

3. LA HAINE

Dans un essai sur la distinction entre faits et valeurs¹, Léo Strauss observe que les historiens et les sociologues ne peuvent pas éviter d'utiliser des termes tels que « cruauté », « vanité », « courage », « cupidité », « dévouement », même lorsqu'ils se font un devoir de respecter strictement une norme de neutralité axiologique, c'est-à-dire d'éliminer tout jugement de valeur dans leurs descriptions et leurs explications. Léo Strauss ajoute : une règle d'élimination de ce genre de termes, manifestement évaluatifs, serait un jeu dont la puérilité ou le cynisme apparaîtrait très facilement dans certains cas extrêmes.

Imaginons, dit-il, qu'un historien se fixe les règles suivantes : il aurait le droit de faire une description purement factuelle des actes accomplis au su et au vu de tous dans un camp de concentration, et une analyse également factuelle des motifs et des mobiles qui ont mû les acteurs, mais il s'interdirait d'utiliser le mot « cruauté ». Une description de ce genre, prétendument directe et objective, ressemblerait, en réalité, ou bien à une satire féroce, ou bien à une euphémisation suspecte ou ridicule.

L'historien ne s'en sortirait pas mieux en choisissant de mettre entre guillemets le terme cruauté. Cette façon de ne pas prendre la responsabilité de l'usage d'un terme évaluatif serait, comme le dit Léo Strauss, un truc enfantin.

À première vue, on pourrait dire de « haine » ce que Léo Strauss dit de « cruauté » : c'est un terme évaluatif et négatif du fait, au moins, qu'il est un antonyme du terme « amour », qu'il serait très difficile de dépouiller de son aspect positif dans notre langage courant, qui est, ainsi que le remarque fort justement Robert Solomon, dans son analyse des passions, un langage humaniste et manichéen².

1. Strauss L., *Droit naturel et histoire* (1954), trad. M. Nathan et É. de Dampierre, Champs, Flammarion, p. 59.

2. Solomon R.C., *The Passions*, Notre Dame, University of Notre Dame Press, 1983.

On pourrait ajouter, à la manière de Léo Strauss, qu'il serait ridicule de recommander de substituer à la phrase « Les nazis haïssaient les juifs » une phrase plus neutre telle que « Les nazis avaient un sentiment d'aversion à l'égard des Juifs et ils ont fait tout ce qui était en leur pouvoir pour les éliminer ».

Cette paraphrase serait incorrecte. D'une part, en substituant au verbe « haïr » la conjonction d'une description d'un état affectif et d'une disposition à se comporter d'une certaine façon, elle supprimerait, ce qui, précisément, caractérise la phrase d'origine : son aspect évaluatif. Dire que « Les nazis haïssaient les juifs », c'est suggérer, au moins, qu'ils avaient à leur égard un comportement ou des sentiments que nous n'apprécions pas.

D'autre part, l'intérêt d'une paraphrase c'est de clarifier une expression confuse. Mais il n'est pas sûr que cette paraphrase soit très éclairante, c'est-à-dire que l'expression « sentiment de répulsion » pose moins de problèmes de compréhension que le terme « haine ». Après tout, « haine » est peut-être l'un de ces termes primitifs indéfinissables et immédiatement intelligibles qui nous servent à comprendre des choses plus compliquées telles que « sentiment de répulsion ».

Cependant, cet exemple ne peut pas suffire à justifier entièrement l'idée que « haïr » et « haine » sont des termes qu'il convient d'aligner sur des termes plus évidemment négatifs tels que « cupidité », « envie », « jalousie », « avarice », « cruauté ».

Mais d'abord, pourquoi ces derniers termes sont-ils si évidemment évaluatifs et négatifs ? C'est parce qu'il semble qu'on ne peut pas leur associer des adjectifs positifs sans dire une absurdité, un peu comme si un trait évaluatif, négatif, était intégré dans la définition même de ces termes et qu'on se contredirait si on le niait.

On ne peut pas, semble-t-il, parler de cupidités « admirables », de jalousies « dignes d'éloges », de malveillances « à montrer en exemple », sans donner l'impression de se contredire ou de ne pas avoir compris ce que signifient cupidité, jalousie, malveillance. La maîtrise de l'usage littéral de ces termes semble impliquer l'exclusion de ce genre d'association, de la même façon que la maîtrise de l'usage du concept « oncle » semble exclure : un oncle sans nièce ni neveu.

Il semble aussi que, de la même façon, une longue tradition philosophique et théologique de condamnation radicale de la

haine jette un doute sur l'acceptabilité de haines « admirables », de haines « à donner en exemple » aux petits enfants, de haines « dignes d'éloges ». En réalité, nous sommes la plupart du temps spinozistes sans le savoir puisque nous pensons généralement « La haine n'est jamais bonne » (ou que si elle est bonne, c'est parce qu'elle est, en vérité, le masque d'un désir de justice ou d'humanité). Mais cela ne semble pas suffire, parce que, manifestement, il nous arrive d'utiliser « haïr » ou « haine » de façon positive, comme lorsque nous disons : « Je hais le mensonge », « Je hais l'égoïsme », c'est-à-dire lorsque l'objet de la haine est mauvais. Par ailleurs, il existe des usages neutres, comme lorsqu'on dit de façon imagée, hyperbolique « Je hais les dimanches ». La question de la valeur morale de ce genre de « haine » ne semble pas se poser.

Mais il me semble que ce sont des usages parasites par rapport à celui sur lequel je voudrais mettre l'accent. Ce qui m'intéresse, ce sont les cas où il semble bien qu'on puisse inférer à partir d'une phrase où figure « haïr » (ou « haine ») l'évaluation « il est haineux ». Or il apparaît vite que cette inférence n'est possible que dans certains cas. De « Il hait les dimanches », par exemple, on ne peut pas inférer « Il est haineux », alors que cette inférence semble toute naturelle de « Il hait les Arabes » à « Il est haineux ».

Je préfère laisser de côté différentes objections qui pourraient affaiblir cette suggestion, car je veux simplement mettre en évidence le fait qu'il serait très difficile de nier qu'il vaut mieux ne pas être jugé haineux (de la même façon qu'il est préférable de ne pas être jugé envieux, aigri, cupide, jaloux, malveillant, vaniteux). Disons, pour conclure cette première discussion, que c'est le caractère ou la disposition qui semble blâmée plutôt que l'acte, et qu'il y a un aspect intrinsèquement évaluatif et négatif dans la haine, lorsqu'elle est envisagée de cette façon.

C'est d'ailleurs en raison de cet aspect intrinsèquement négatif du terme haine que certaines précautions devraient être prises, à mon sens, dans l'attribution de la haine à autrui. Dire de quelqu'un qu'il est haineux ce n'est certainement pas lui faire un compliment. Et il y aurait beaucoup de choses à dire sur cette manie qu'ont les journalistes, par exemple, d'imputer la haine à certaines personnes ou classes de personnes, assez généralement des personnes ou des classes de personnes dont le statut est bas

(les pauvres, les primitifs, les gens de l'Est, etc.). On peut se demander, bien sûr, s'ils imputeraient avec la même insouciance la haine à des personnes ou des classes de personnes plus proches ou plus appréciées socialement. Même si les comportements étaient identiques, ils parleraient probablement de comportements implacables, volontaires, décidés, violents à la rigueur, mais pas haineux.

Quoi qu'il en soit, l'ensemble de ces considérations que je ne peux pas approfondir ici¹ devrait suffire, pour le moment, à donner raison, à première vue, à Léo Strauss ou, du moins, à approuver deux de ses thèses :

1) Il y a des termes intrinsèquement évaluatifs.

2) Ces termes sont inéliminables dans nos façons de parler du comportement humain.

Je voudrais souligner seulement que ces deux thèses sont loin d'être des platitudes ou des vérités d'évidence. La première s'oppose à des théories sémantiques solidement établies. Les deux prises ensemble menacent sérieusement l'analyse causale des conduites humaines, ce qui est particulièrement intéressant dans le cas qui me préoccupe.

En effet, la haine sert d'explication causale passe-partout. Elle explique le comportement du corbeau de la Vologne, des Serbes, des Croates, des néo-nazis et ainsi de suite. Je voudrais montrer que ce genre d'explication causale est incompatible avec le fait que haine est un terme évaluatif. Bref, je voudrais essayer d'atteindre le préjugé causal, particulièrement tenace dans le cas de la haine.

Mais je vais revenir d'abord sur le problème des termes intrinsèquement évaluatifs, c'est-à-dire sur la difficulté sémantique, dont Léo Strauss n'a pas tenu compte, semble-t-il.

1

En affirmant que « haine », ou « cupidité », « jalousie », « envie », « malveillance » sont des termes intrinsèquement évaluatifs, on se range dans un camp sémantique bien précis.

1. Je me permets de renvoyer à mon *Portrait logique et moral de la haine*, Combas, L'éclat, 1993.

Il fut un temps où la tendance était à montrer que certaines de nos descriptions étaient, en réalité, des expressions d'émotions ou des prescriptions déguisées. Il semblait clair qu'un énoncé axiologique tel que « Les hommes sont des canailles » ne décrivait rien, mais exprimait seulement un sentiment négatif sur les hommes ou une recommandation (de ne pas trop les fréquenter, par exemple). Il n'a pas été très difficile d'étendre ces théories du sens et de la vérité à des cas moins évidents tels que « J'ai mal aux dents », « La terre est ronde ». Chaque énoncé apparemment descriptif avait un sens non descriptif particulier. Dans cette perspective, lorsque je dis : « J'ai mal aux dents », je ne décris rien, je demande de l'aide ; lorsque je dis « La terre est ronde », je ne décris pas la terre, j'affirme quelque chose de la terre, relativement à certaines croyances qui ne sont pas en l'occurrence celles de la physique, laquelle exige beaucoup plus de précision. Mais un argument inspiré de Frege a été proposé contre cette théorie. Cet argument nous recommande de distinguer très soigneusement ce que dit un énoncé, ou ce que signifie un terme de l'acte que nous accomplissons en l'utilisant (affirmer, nier, prescrire, blâmer, etc.), c'est-à-dire d'adopter à l'égard du terme ou de l'énoncé une attitude neutre comme lorsque nous nous contentons de le citer, de le mentionner, sans le prendre à notre compte. Il y a une différence entre affirmer que « La terre est plate » et supposer que « La terre est plate », mais cela ne veut pas dire que l'expression « La terre est plate » ait deux sens différents : le premier affirmatif, le second conditionnel. De même, il y a une différence entre affirmer que « mentir est un mal » et supposer que si mentir est un mal, pousser son petit frère à mentir est un mal, mais cela ne veut pas dire que l'expression « mentir est un mal » ait deux sens différents. Le premier prescriptif, le second non prescriptif.

Les théories émotivistes ou prescriptivistes nous obligent à multiplier les significations sans nécessité ou plus exactement à faire proliférer indéfiniment les classes d'énoncés. Il vaut peut-être mieux, comme on l'a vu à propos des théories pragmatiques de la signification, s'en tenir à un principe plus économique, qu'on peut appeler moniste, principe qu'on peut respecter en distinguant deux niveaux : celui du sens des mots, donné par ses conditions de vérité, qu'on doit essayer de dégager indépendam-

ment de tout contenu évaluatif, prescriptif ou autre et celui des actes que nous accomplissons en les utilisant. Bref, de distinguer la signification et l'emploi, et de rejeter l'idée qu'il y aurait des termes intrinsèquement évaluatifs. Bernard Williams nous a recommandé d'appliquer ce principe au langage moral en distinguant les termes généraux tels que « bien » et « mal » des termes spécifiques tels que « cruauté », « haine », « courage », « jalousie ». Pour les termes spécifiques qu'il appelle aussi termes épais, la mise en évidence d'un niveau purement descriptif ne pose, d'après lui, aucun problème. Quant aux termes qu'il appelle généraux, il estime qu'on a tout intérêt à les éliminer¹.

Mais le destin des théories en philosophie du langage est d'être réfutées ou contestées et c'est bien ce qui se produit actuellement. Certains philosophes du langage rejettent l'idée qu'il serait possible d'isoler une composante purement descriptive dans les énoncés, en général, ou même de dégager quelque chose qui ressemblerait à une proposition parfaitement épurée, susceptible d'être évaluée en termes de vérité ou de fausseté, indépendamment de tout contexte. Bref, la thèse des expressions intrinsèquement évaluatives n'est pas définitivement enterrée, mais, elle doit être reconstruite contre les arguments dits frégeïens². Je vais laisser là ces difficultés sémantiques pour passer à mon second argument. Mais je tenais à souligner que les affirmations de Léo Strauss demandaient un sérieux examen. Elles ne sont pas indéfendables, mais elles doivent être mieux défendues.

2

Lorsque nous parlons de la haine nous sommes souvent victimes d'un puissant préjugé causal. Nous expliquons certains comportements en disant : c'est à cause de la haine. Puis nous essayons d'expliquer la haine en invoquant différents facteurs psychologiques ou sociologiques, tels que la frustration ou l'oppression mêlée à la dépendance. Parfois, nous nous contentons de donner un rôle fonctionnel à la haine, mais c'est parce que nous

1. Williams B., *L'Éthique et les limites de la philosophie*, Paris, Gallimard, p. 132-143 plus spécialement p. 141-142. Cf. aussi note 91.

2. Cf., entre autres, le compte rendu de F. Récanati, *op. cit.*, 1991.

la considérons comme une cause non causée, un attribut naturel, consubstantiel, de l'homme qui pourrait dans une certaine mesure le servir et, en certains cas, le desservir. L'homme aurait, par nature, une tendance à la destruction, la violence, l'agression, la répulsion à l'égard d'autrui. Il serait naturellement haineux. Je passe sur différents raffinements de ces théories, tels que la thèse disant que la seule haine naturelle, consubstantielle, serait la haine de soi, le désir d'auto-destruction, la haine d'autrui n'étant qu'une projection, une déviation du but initial, une construction psychologique ou sociale. Dans tous les cas, nous concevons la haine comme une sorte de phénomène naturel pris dans l'immense réseau des causes, c'est-à-dire comme effet d'une cause et cause à son tour d'un effet : un effet pénible, généralement.

Cependant, pour traiter la haine en tant que cause ou effet dans l'ordre naturel, nous avons besoin d'avoir une sorte de définition opératoire. Nous devons être en mesure, par exemple, d'identifier la haine relativement à d'autres sentiments et comportements, car si tel n'était pas le cas nos tentatives de l'expliquer (sociologiquement, psychologiquement, biologiquement) deviendraient futiles. De même, les explications par la haine seront creuses si nous ne réussissons pas à identifier correctement ce qu'elle est. Mais comment faisons-nous pour distinguer la haine de la violence, de l'envie, de la jalousie, de l'aversion, de la cruauté, de la répulsion, de la rancœur, du ressentiment, de l'indifférence, de l'absence de pitié ? Nous avons à notre disposition trois principes descriptifs :

1) La haine est un état qualitatif, quelque chose qu'un sujet ressent, éprouve. La haine est un affect, un sentiment.

2) La haine est un état intentionnel, une certaine façon de viser des objets, de se situer dans le monde.

3) La haine est un comportement ou une disposition à se comporter d'une certaine façon, disons agressive ou hostile.

Remarquons que la haine en tant qu'état intentionnel reçoit une définition plutôt négative. En fait, elle est définie comme n'étant ni un état qualitatif ni un pur comportement, mais une sorte de relation à des objets, qui en même temps la définissent. C'est un point de vue qu'il n'est pas très facile de comprendre. J'espère qu'il s'éclairera lorsque je donnerai, tout à l'heure, un exemple.

Aucun de ces principes ne suffit, à mon sens, à individuer la haine ou à l'individuer assez pour qu'elle puisse nous apparaître comme un prétendant acceptable à l'explication causale.

Commençons par la définition qualitative ou affective de la haine. Pour qu'elle soit acceptable il faudrait que nous ayons un critère de discrimination qualitatif assez fin qui nous permettrait de distinguer la haine de l'envie, de la jalousie, du ressentiment, de l'aversion, de la rancœur, de l'agressivité et même de l'indifférence.

Je crois que nous n'avons pas la moindre idée de ce que pourrait être une description fine de ces différences en termes qualitatifs. Je suis persuadé aussi que nos meilleurs romanciers ou nos plus profonds poètes seraient bien incapables de nous donner des conseils à ce sujet, pour des raisons qui n'ont rien d'accidentel.

En examinant la théorie de l'orgueil de David Hume, Davidson faisait la remarque suivante : « On pense souvent que, selon Hume, être fier d'être habile (par exemple), c'est simplement faire l'expérience d'une certaine "qualité ultime ressentie". Cette conception serait sûrement fausse parce qu'elle ne nous donnerait aucune possibilité de distinguer entre le fait d'être fier d'être habile et celui d'être fier d'être gentil avec les Kangourous¹. »

Il me semble que cette remarque est très juste et qu'on peut, en l'exploitant, l'appliquer à nos façons d'identifier nos prétendus états qualitatifs. Ce qui nous permet de distinguer la jalousie de l'envie, par exemple, ce n'est certainement pas une certaine « qualité ultime ressentie » mais plutôt une certaine façon de se comporter ou de se situer par rapport à certains objets. Dans la jalousie, par exemple, nous souhaitons conserver quelque chose alors que dans l'envie nous voulons plutôt nous approprier quelque chose. Est-ce que ces relations différentes à l'objet correspondent à des états qualitatifs différents ? Ce n'est pas nécessaire. On peut envisager la possibilité qu'à deux sentiments différents (l'envie et la jalousie, par exemple) correspond un *même* état qualitatif (une sorte de douleur) et, inversement, qu'à un même sentiment (la jalousie, encore) correspondent *plusieurs* états qualitatifs diffé-

1. Davidson D., *op. cit.*, 1993, p. 366.

rents (des douleurs allant du côté de l'excitation ou du côté de l'abattement), tout ce qui est dit de ces états qualitatifs restant très vague, bien entendu.

Passons donc aux deux autres grands principes, l'intentionnel et le comportemental, ce qui nous donnera l'occasion de faire deux petites excursions dans le XVIII^e siècle puis dans l'Antiquité. D'abord la haine intentionnelle et le XVIII^e siècle.

3

Dans son analyse des passions, Hume avait isolé ce qu'il appelait des passions indirectes, c'est-à-dire des impressions plaisantes ou douloureuses liées à des croyances, à des idées¹.

À l'impression directe et douloureuse causée par l'impact du poing sur le nez peut succéder la passion de la honte si, entre cette impression et cette passion, s'est interposée, par exemple, l'idée qu'on a été lâche en ne frappant pas celui qui nous a frappé. Et c'est pourquoi la honte est une passion indirecte.

Davidson a subtilement montré qu'en réalité Hume avait défendu une théorie propositionnelle des passions indirectes, celles-ci n'étant pas individuées ou définies par un certain état qualitatif du sujet, mais par une description de l'idée qui les causent, laquelle a la forme d'une proposition dans notre langage². Davidson a donc aligné l'analyse des passions indirectes sur celle, plus générale, des croyances. Une croyance n'est pas un état affectif ou qualitatif du sujet, mais une attitude ou un jugement d'approbation à l'égard d'une proposition particulière. Il paraît assez évident que croire que deux et deux font quatre, ce n'est pas être dans un état qualitatif particulier ou, plus exactement, cet état n'est pas pertinent pour définir ou identifier la croyance (on peut être très triste ou très joyeux si deux et deux font quatre, mais cela n'a aucune importance si ce qu'on veut, c'est définir ou identifier cette croyance).

1. Hume D., *Les passions* (Livre II du *Traité de la Nature Humaine*), trad. fr. J.P. Cléro, Paris, Garnier-Flammarion, p. 110.

2. Davidson D., *op. cit.*, 1993, p. 365-383.

L'avantage de ce genre d'analyse c'est qu'elle semble fournir un bon critère d'individuation d'un état mental, ce qui nous autorise à lui donner la place d'un antécédent acceptable dans une explication de type causal.

Ainsi, je peux spécifier le contenu de ma croyance en disant que c'est la croyance qu'il y a du sel sur l'étagère. Et c'est ce contenu, c'est cette croyance bien individuée ou spécifiée par la proposition « il y a du sel sur l'étagère » qui peut servir à expliquer causalement pourquoi je tends ma main vers l'étagère lorsque je veux saler mes spaghettis.

Ce genre d'analyse présente deux difficultés. La première est assez spécifique; elle pourrait suffire à exclure la possibilité d'une analyse intentionnelle ou propositionnelle de la haine. Mais j'insisterai surtout sur la seconde, dont les implications théoriques sont beaucoup plus intéressantes. Voyons d'abord la difficulté spécifique.

L'analyse propositionnelle ne peut pas s'appliquer à la haine, tout simplement parce que le verbe « haïr » réclame avant tout des compléments nominaux. On hait quelque chose ou quelqu'un plutôt que quelque chose ou quelqu'un sous une description. La haine est une relation directe à une personne et non une relation à une proposition.

Et cette caractéristique grammaticale correspond assez bien à l'idée que nous nous faisons de ce sentiment. On est censé avoir la possibilité de haïr quelqu'un sous toutes ses descriptions, c'est-à-dire, en dépit de tout ce qu'on sait ou apprendra à son propos, de même qu'on est censé avoir la possibilité d'aimer quelqu'un sous toutes ses descriptions, c'est-à-dire en dépit de tout ce qu'on sait ou apprendra à son propos.

Or si la haine n'a pas de complément propositionnel, elle ne peut pas être individuée sur le modèle de l'individuation des croyances ; elle reste aussi mal définie en tant qu'état intentionnel qu'en tant qu'état qualitatif, et on ne peut pas lui attribuer un rôle causal spécifique.

Voyons, à présent, pourquoi la difficulté générale est plus intéressante. Dire que le contenu d'une croyance est donné par une proposition et que cela s'exprime dans le fait que c'est en vertu de ce contenu que les croyances ont une efficacité causale, c'est adopter une attitude qu'on peut appeler internaliste relative-

ment à l'identification des croyances. En effet, défendre cette position revient à dire que les croyances sont intentionnelles ou inexistantes ou n'existent que dans l'esprit du sujet, dans la mesure où l'existence de l'état de choses qu'elles visent n'a pas de pertinence sémantique. Si je crois qu'il y a du sel sur l'étagère, cette croyance est indépendante de l'état de choses qui est décrit dans la proposition qui suit l'expression de ma croyance. Il peut être vrai que je crois qu'il y a du sel sur l'étagère, quel que soit l'état réel du monde, c'est-à-dire, même s'il n'y a pas de sel sur l'étagère. Et c'est cette croyance purement interne, indifférente à l'existence effective de son objet qui cause mon comportement et, aussi, en l'occurrence ma déception puisqu'il n'y a pas de sel sur l'étagère.

À cette analyse dite solipsiste ou internaliste, on oppose généralement deux arguments:

1) l'argument des pensées singulières.

2) l'argument du holisme sémantique.

1) Certaines pensées ne peuvent avoir un sens déterminé que si celui qui a ces pensées est en contact direct ou indirect avec un objet du monde. C'est le cas des pensées indexicales, celles qui introduisent un terme démonstratif désignant un objet du monde ou celles qui font intervenir un individu désigné par un nom propre. Dans ces cas, la croyance n'est pas indifférente au monde. Elle est définie de façon externaliste, comme on dit.

On peut ajouter que les arguments de Putnam sont, eux aussi, typiquement externalistes, en ce sens¹. Putnam pense qu'à deux états mentaux parfaitement identiques ou indiscernables peuvent correspondre des croyances différentes, selon le chemin qu'elles ont pris pour devenir nos croyances. Il pourrait se faire qu'un cerveau dans une cuve soit exactement dans le même état mental que moi lorsque je pense « Il y a du sel sur l'étagère », mais nous n'aurons pas les mêmes croyances, car l'histoire de ma relation au sel et aux étagères est différente de la sienne. Bref, pour les externalistes il est absurde de dire, à la manière cartésienne des internalistes, que nos croyances pourraient être les mêmes, même si le monde n'existait pas ou s'il était entièrement différent.

1. Putnam H., à différents endroits. En français, dans *Raison, vérité et histoire*, trad. fr. A. Gershenfeld, Paris, Minuit, 1989.

La grande difficulté, pour les externalistes, c'est d'expliquer comment ces croyances qui, d'une certaine façon, ne sont pas entièrement dans la tête des gens peuvent les faire agir, ou comment l'externalisme est compatible avec le causalisme. En règle générale, on estime que l'externalisme n'est pas compatible avec l'explication causale. C'est un argument dont je me servirai. Je ne pourrai l'illustrer ici qu'au moyen d'un exemple très imparfait.

Supposons que je possède un tableau que je crois être un faux Vermeer. En réalité, c'est un authentique Vermeer, ainsi que pourraient l'attester des experts, qui établiraient un lien causal entre ce tableau et la main de Vermeer. L'avis des experts, la relation causale de la main de Vermeer à ce tableau seraient des éléments externes relativement à ce qui se passe dans mon esprit. Mais ce qui pourrait expliquer mon comportement ne pourrait absolument pas tenir compte de ces identifications externes. J'agisrais en fonction de ma croyance que c'est un faux Vermeer¹.

2) De même que le sens d'un mot vient de la place qu'il occupe dans une phrase, de même le sens d'une proposition vient de la place qu'elle occupe dans le système entier de nos croyances.

Nos croyances ne se présentent pas une à une devant le tribunal de l'expérience. Elles forment un système enchevêtré. Une croyance aussi banale que « Il y a du sel sur l'étagère » n'a de sens que du fait qu'elle est liée à une infinité d'autres croyances portant sur ce qu'est le sel ou une étagère. De sorte qu'il n'est pas du tout certain que la référence à une proposition suffise à spécifier, à individuer correctement un état mental, afin de lui faire jouer un rôle causal.

4

Nous pouvons passer à présent à l'identification par le comportement — et à l'Antiquité.

1. Ceux qui veulent avoir des détails sur le fonctionnement de l'argument ou les subtilités introduites par les internalistes qui distinguent le contenu étroit (interne) et large (externe) des croyances, et font du contenu large un phénomène asymétriquement dépendant ou un épiphénomène, pourront lire Pascal Engel, *op. cit.*, 1994c, p. 550-556. Pierre Jacob, *op. cit.*, 1993, p. 391-413.

Lorsque Plutarque essaie d'identifier ce qui distingue la haine de l'envie, de même que lorsqu'Aristote veut nous donner les moyens de distinguer la haine et la colère, ils font référence à des styles de comportements plutôt qu'à des états qualitatifs ou mentaux¹. Toutefois, cette orientation comportementaliste est plus clairement marquée chez Plutarque.

D'après lui, la haine et l'envie ont quelque chose en commun. Dans la haine comme dans l'envie, il nous arrive de dénigrer systématiquement autrui, de nous réjouir de ses échecs, de nous affliger de ses succès. Cependant Plutarque estime qu'il y a une différence dans ce qu'il appelle la « croissance » de ces sentiments. Il est difficile de résister à la tentation de traduire l'analyse de Plutarque en termes quasi quantitatifs. Selon Plutarque, la courbe d'évolution de la haine n'a pas la même allure que la courbe d'évolution de l'envie. Dans l'envie, notre comportement hostile dure tant que la personne enviée occupe une position approximativement équivalente à la nôtre. Mais ce comportement hostile cesse souvent lorsque la position de la personne enviée s'écarte considérablement de la nôtre. Selon Plutarque, « Le succès, dans sa grandeur et son éclat suprêmes, souvent éteint l'envie. Il n'est pas vraisemblable en effet qu'on ait pu envier Alexandre ou Cyrus quand ils eurent triomphé et furent devenus maîtres du monde². » Par ailleurs, l'envie disparaît complètement lorsque la personne enviée est frappée par un malheur : les échecs font cesser l'envie.

Dans la haine, il en va tout autrement. La haine ne se relâche pas devant la supériorité flagrante de l'ennemi ou sa puissance. Quant aux malheurs et aux échecs de l'ennemi, ils ne font pas cesser la haine. Plutarque estime qu'on peut continuer de haïr ses ennemis, même lorsqu'ils sont vaincus ou humiliés. Plutarque conclut : la haine n'abandonne sa victime ni dans le bonheur ni dans le malheur, tandis que l'envie renonce devant l'excès de l'un ou de l'autre.

1 Aristote, « De l'amitié et de la haine », *Rhétorique* II, éd. et trad. M. Dufour, Paris, Les Belles Lettres, 1967, p. 68-72 ; Plutarque, « De l'envie et de la haine », *Œuvres Morales*, Tome VII, deuxième partie, éd. et trad. R. Klaerr et Y. Vernière, Paris, Les Belles Lettres, 1974, p. 50-55.

2. Plutarque, *op. cit.*, 1974, p. 53.

En fait, Plutarque nous propose une sorte de protocole expérimental qui devrait nous permettre de distinguer objectivement la haine et l'envie. Si notre comportement hostile se maintient, même lorsque la personne visée nous a largement dépassé par ses succès ou lorsqu'il lui arrive un grand malheur, c'est que nous sommes haineux. Si, au contraire, nous abandonnons notre comportement hostile à la vue des succès flagrants de la personne visée ou lorsqu'elle souffre, c'est que nous étions seulement envieux.

Il n'y a rien d'embarrassant dans cette façon de définir un prétendu sentiment par le comportement, plutôt que par un certain état qualitatif ou mental.

Comme Ryle l'a observé, lorsque quelqu'un est vaniteux, il est souvent le dernier à s'en apercevoir¹. Si c'était un état qualitatif, il serait le premier. On pourrait dire qu'il en va de même pour la haine et l'envie ; c'est en adoptant un point de vue objectif sur soi-même qu'on peut s'apercevoir qu'on est haineux ou envieux, et cela, seulement, si on a, à sa disposition, une procédure semblable à celle que Plutarque propose.

Cependant, l'identification comportementale souffre des défauts évidents de toutes les constructions behavioristes. Il reste toujours des termes mentaux inanalysés qui traînent dans les descriptions (le sentiment d'hostilité par exemple) ; et certaines attitudes ne s'expriment pas dans des actes, ce qui rend vaine toute identification comportementale. Il y a aussi le problème plus général de l'interprétation.

Il n'est pas nécessaire d'être un Nietzsche obsessionnel pour démasquer, derrière la plupart de nos comportements altruistes, de la méchanceté, de la vanité, de l'égoïsme et, à l'inverse, il n'est pas impossible de prétendre qu'un comportement brutal et violent exprime un amour authentique. Bref, le soupçon que nos actes n'expriment pas vraiment nos intentions étant indéfectible, le critère de l'identification comportementale paraît toujours grossier.

Toutefois, on peut avoir des raisons de penser qu'il y a des limites à l'interprétation. Le contexte nous permettrait d'exclure celles qui sont inacceptables. Mais faire intervenir le contexte

1. Ryle G., *op. cit.*, 1978, p. 85.

dans la procédure d'identification de la haine, c'est ajouter à la difficulté : quels sont les critères d'identification du contexte ? Ses conditions d'identité ? D'autre part, c'est introduire une définition externaliste de l'acte ou du sentiment, ce qui annule la possibilité de proposer ensuite une explication causale, selon le principe d'incompatibilité entre externalisme et causalisme que j'ai évoqué.

5

Ce qu'on peut retenir de cette analyse, c'est que le sujet haineux a tendance à ne pas s'attribuer un caractère ou des dispositions haineuses, tout simplement peut-être, parce qu'il n'a guère envie de s'attribuer un caractère ou une disposition socialement dévalorisée, moralement dévaluée. Mais cela ne signifie pas qu'il a ces sentiments ou ces comportements avant l'évaluation morale.

Tout ce que nous pouvons dire, c'est que les états de haine sont, semble-t-il, attribués dans le cours d'une sorte de procédure objective, à la troisième personne, une procédure que le sujet dit haineux peut conduire lui-même en réfléchissant sur ses actions.

Ce n'est pas en tant qu'état purement affectif, purement intentionnel, purement comportemental que la haine est attribuée et identifiée. C'est dans une combinaison variable des trois, à laquelle se mêle inévitablement un jugement moral à la troisième personne.

De sorte que la haine apparaît en définitive comme une sorte d'état moral. Pour parodier une formule fameuse je dirai : on n'est pas haineux, mais on le devient lorsque des états qualitatifs, des intentions, des comportements font l'objet d'un jugement moral négatif.

Il en résulte qu'on peut envisager la possibilité que deux comportements ou styles de comportements objectivement indiscernables ou deux états qualitatifs identiques seront distingués par le fait qu'ils se présentent dans des contextes moraux différents (on peut penser, même si l'exemple est usé, à la différence entre la violence en temps de guerre et aux mêmes actes accomplis en temps de paix).

Tout cela me permet d'établir un lien entre l'argument comportemental et l'argument intentionnel et de préciser ma critique du préjugé causal, pour conclure.

Nous avons vu qu'il était possible de donner une définition externaliste des croyances, mais qu'elle avait l'inconvénient de bloquer l'analyse causale de l'efficacité des croyances. Si les croyances se présentent à nous en bloc, ou si elles se définissent autrement que par ce qui se passe dans la tête des gens, elles n'ont pas de pertinence causale relativement au comportement humain.

Or ces identifications de la haine à la troisième personne, en tant qu'attribut moral, sont typiquement externalistes. Elles ne dépendent pas de ce qui se passe dans l'âme du sujet dit haineux. Et, comme les autres analyses externalistes, elles sont incompatibles avec l'analyse causale, telle qu'on l'entend habituellement.

En résumé : si la haine est une attribution morale à la troisième personne, comme il semble bien que ce soit le cas, il est difficile de voir comment on peut la faire figurer dans une explication causale ordinaire, internaliste en principe.

Je suis conscient de la fragilité de mon argument. En réalité, c'est une piste de recherche seulement, qui repose sur l'incompatibilité présumée (je dis bien présumée) entre causalisme et externalisme.

Mais on peut se contenter de dire que si l'analyse causale de la haine est viciée, c'est tout simplement, parce que nous n'avons aucun bon moyen d'établir les conditions d'identité de la haine, de l'identifier par opposition, au moins, à d'autres sentiments voisins. Ni le critère qualitatif, ni le critère intentionnel, ni le critère comportemental ne sont satisfaisants. Quant au critère moral, il exclut, semble-t-il, une analyse causale ordinaire.

TROISIÈME PARTIE

L'AUTONOMIE DE L'ÉTHIQUE

1. RÉALISME OU RELATIVISME : QUOI DE NEUF ?

À quelques exceptions près, les sociologues et les historiens n'ont jamais hésité à défendre l'objectivité des valeurs et des normes. Ce qui leur fait dire que les normes et les valeurs sont objectives, c'est qu'elles ont, d'après eux, des propriétés de ce genre.

1) Nous ne pouvons pas les modifier volontairement alors que cette possibilité n'est jamais exclue pour des lois ou des règles de jeu.

2) Nous savons très bien les distinguer de nos préférences ou de nos désirs.

3) Elles sont là avant nous et elles continuent d'exister après nous. (On peut envisager une version plutôt concrète ou historique : on les trouve toutes faites à notre naissance et elles continuent d'exister après notre disparition ; et une version plus abstraite ou générale : il n'est pas inconcevable que les normes et les valeurs restent approximativement les mêmes, alors que les individus qui s'y conforment ou les violent, sont différents, d'où l'idée que les normes et les valeurs préexistent aux individus ou existent indépendamment d'eux, c'est-à-dire objectivement.)

Il existe bien sûr d'autres raisons, moins « durkheimiennes », de prêter une sorte d'objectivité aux valeurs et aux normes, mais elles ne sont pas dépourvues de toute relation avec celles-ci.

Cependant, une chose est de dire que les valeurs et les normes sont, en ces différents sens, « en dehors de nous »¹, et qu'elles exercent « sur nous » une sorte de contrainte, autre chose est d'affirmer que ces valeurs ou ces normes pourraient être vraies ou fausses au sens le plus fort du terme, c'est-à-dire de la même manière que les propositions mathématiques ou empiriques.

Il est facile de comprendre : « Il est soit vrai soit faux que les Français sont, en majorité, pour la peine de mort » (même si on

1. Durkheim E., « Jugements de valeur et jugements de réalité » (1911), *Sociologie et philosophie*, éd. par C. Bouglé, Paris, PUF, 1974, p. 102-121.

juge, en définitive, que la proposition est indécidable parce qu'il n'existe aucun bon moyen de savoir ce que sont exactement les « croyances » ou les « attitudes » des « Français » et ce que signifie exactement l'affirmation publique qu'on est « pour » ou « contre » quelque chose).

On n'aura aucune difficulté non plus à comprendre : « Il est soit vrai soit faux que la peine de mort est, en France, un châtement légal » (même si on estime que, tout compte fait, la proposition est indécidable parce que le concept de « loi » est indéfinissable).

La situation est un peu différente lorsqu'on est confronté à des énoncés tels que : « Il est soit vrai soit faux que la peine de mort est un mal », les corollaires étant, si la proposition est vraie :

1) les Français qui croient que la peine de mort est un bien se *trompent* exactement de la même façon que les philosophes de l'antiquité qui croyaient que la terre était plate ou immobile ;

2) la loi instituant la peine de mort est injuste, dans la mesure où une loi juste ne peut pas être incompatible avec des principes moraux vrais.

Il me semble qu'on aurait du mal à trouver des historiens ou des sociologues qui, en tant que quels, accepteraient de donner une sorte de signification empirique à ces dernières propositions. Ces disciplines exigent, semble-t-il, une certaine dose de relativisme, fut-elle presque imperceptible¹. Durkheim, qui n'était pas particulièrement modéré en matière d'objectivité des valeurs et des normes, n'a jamais suggéré qu'il pourrait exister une réalité morale éternelle, laquelle serait susceptible de rendre vraies ou fausses des propositions telles que « La peine de mort est injuste », « L'esclavage est un mal », etc.²

D'un autre côté, il y a toujours eu des philosophes pour défendre avec ardeur l'universalité des valeurs et des normes ou, plus exactement, la vérité de certaines valeurs et de certaines normes,

1. C'est l'idée que Guy Rocher, par exemple, défend dans son ouvrage d'introduction à la sociologie générale : *L'Action sociale*, Paris, Points-Seuil, 1968, p. 57. Mais c'est peut-être un point de vue qui est en train de changer. Cf. Raymond Boudon, *Le Juste et le Vrai. Études sur l'objectivité des valeurs et de la connaissance*, Paris, Fayard, 1995.

2. Durkheim E., *op. cit.*, 1974, p. 112.

vérité qui échapperait hélas, à la plupart de ceux qui les lisent. Ils ont réussi, à la rigueur, à sauver l'idée que le comportement humain n'était pas nécessairement conformiste et qu'il fallait bien rendre compte de la capacité que nous avons de critiquer les conventions existantes, capacité dont nous savons depuis toujours qu'elle existe (l'exemple d'Antigone peut être servi pour persuader ceux qui n'en seraient pas convaincus). Mais ils n'ont certainement pas réussi à faire admettre que cette capacité critique témoignait d'elle-même de l'existence de vérités éthiques transcendantes.

Il faut bien reconnaître qu'il n'est pas très facile de traduire ce genre d'argument, de telle sorte qu'il puisse faire l'objet d'un débat authentique avec les spécialistes des disciplines relativistes (ou pour ne pas fermer d'emblée toutes les possibilités, avec les relativistes dans ces disciplines). Même un argument épistémologique modeste tel que : la meilleure explication possible de la force de la critique morale, c'est qu'elle est *vraie* leur paraîtra probablement suspect.

Je trouve donc particulièrement réjouissant qu'une discussion de très vaste portée se soit engagée, parmi les philosophes analytiques, autour de la question du réalisme moral¹, qui est celle de la possibilité pour les énoncés moraux d'être réputés vrais ou faux. Cette discussion ne vise pas à en finir une fois pour toutes avec le relativisme ou la dichotomie du fait et de la valeur. Son ambition est plus modeste (et plus « réaliste » pour utiliser le mot dans un autre sens) : donner une certaine plausibilité aux arguments qui plaident en faveur de la possibilité pour les énoncés moraux d'être vrais ou faux et pas seulement « acceptables dans un certain contexte », c'est-à-dire rendre ces arguments légitimes même aux yeux de ceux qui font de la croyance au relativisme ou à la dichotomie du fait et de la valeur un véritable article de foi.

Personnellement, je défendrai l'idée que ces arguments ne sont pas décisifs. Ma stratégie consistera à essayer de montrer que les différentes espèces de réalisme moral se détruisent réciproquement. C'est donc une thèse plutôt négative que je soutiendrai sur ce sujet.

1. Cf. Virvidakis S., « Variétés du réalisme en philosophie morale », *Philosophie*, 22, 1989, p. 11-35 ; Darwall S., Gibbard A., Railton P., « Toward Fin de Siècle Ethics : Some Trends », *The Philosophical Review*, 101, 1, 1992, p. 115-189.

Mais j'essaierai de souligner, aussi, l'importance de la contribution (négative également) des arguments réalistes à la philosophie morale et aux sciences humaines en général. Grâce à eux, nous avons appris que la plupart des arguments qu'on utilise habituellement pour justifier le dualisme du fait et de la valeur, la distinction évaluatif-descriptif, le relativisme en général ne sont absolument pas concluants. Plus spécifiquement, nous avons appris que nier la possibilité de l'objectivité morale pourrait nous obliger à nier toute forme d'objectivité, car nos énoncés dits empiriques ne résisteraient pas à certains tests qu'on fait subir principalement aux énoncés moraux afin de prouver qu'ils ne sont pas susceptibles d'être vrais ou faux. Et, inversement, nous avons appris que les énoncés moraux étaient tout à fait susceptibles de subir avec succès d'autres tests, relatifs à d'autres théories dont nous nous servons aussi pour décider de la possibilité pour un énoncé d'être vrai ou faux.

Pourtant, du fait que ce qu'on appelle la « révolution » réaliste¹ a disqualifié la plupart des arguments traditionnels en faveur de la dichotomie du fait et de la valeur ou du relativisme, on ne peut pas conclure que cette dichotomie et cette thèse sont absurdes, stériles ou fausses.

Tout ce que nous avons perdu, jusqu'à présent, c'est un certain genre de justification de ces idées. Et tout ce que la querelle du réalisme moral peut avoir engendré, c'est la nécessité de trouver des arguments nouveaux et plus convaincants (ce qui n'est vraiment pas un mal, soit dit en passant). Car il serait absurde de penser que le conflit entre les partisans et les adversaires de la dichotomie fait-valeur et du relativisme pourrait être tranché de manière définitive par un de ces arguments décisifs que tout le monde cherche et que personne ne trouve, pour la bonne raison qu'il n'y en a pas.

1

Russell disait : « Je ne vois pas très bien comment réfuter les arguments qui plaident en faveur de la subjectivité des valeurs

1. C'est Jonathan Dancy qui parle d'une « révolution » affectant tous les secteurs de la philosophie morale, dans « Transatlantic Variants », *T.L.S.*, 23-29 mars 1990, p. 326.

éthiques, mais je suis parfaitement incapable de croire que tout ce qu'il y a de mal dans la cruauté délibérée, c'est que je ne l'apprécie pas¹».

Après la « révolution » réaliste, c'est plutôt l'inverse qu'on pourrait avoir des raisons de vouloir affirmer : « Je ne vois pas très bien comment réfuter les arguments qui plaident en faveur de la réalité objective des valeurs éthiques, mais je suis parfaitement incapable de croire qu'il n'y a aucune différence du point de vue de l'objectivité entre "La neige est blanche", "Deux et deux font quatre", "L'argile est imperméable", "Ce pont est dangereux", "Cette histoire est drôle" et "Les vaches sont sacrées", "Mère Teresa est bonne", "Il faut tenir ses promesses" ».

Comment les réalistes moraux ont-ils réussi à rendre des phrases équivalentes ? De quatre manières différentes, au moins, c'est-à-dire en construisant quatre variétés de réalisme moral² au moins, qui pourraient être baptisées de la façon suivante : sémantique, ontologique, interne, normatif³. Voici leurs principaux caractères.

1. Russell B., « Notes on Philosophy », *Philosophy*, 35, 1960, p.146-147.

2. J'emprunte cette heureuse expression à Stelios Virvidakis, *op. cit.*, 1989.

3. Les deux premiers « réalismes » sont des théories « diffuses » : elles n'ont ni appellation spécifique ni défenseur attitré. J'ai appelé « réalisme sémantique » celui qui est défendu, entre autres, par Mark Platts dans « Moral Reality », dans Sayre-Mc Cord G., éd., *Essays on Moral Realism*, Ithaca, Cornell University Press, 1988, p. 282-300. J'ai hésité à propos du nom qu'il convenait d'utiliser pour baptiser le second. Réalisme « axiologique » aurait peut-être été préférable à « ontologique ». Mais je ne voulais pas exclure d'emblée les normes du tableau ontologique. Je comptais même développer, à partir de la difficulté de trouver une bonne ontologie des normes, un argument contre le réalisme axiologique. Je ne suis pas allé très loin dans cette direction, mais suffisamment tout de même pour préférer un terme moins spécifique que « axiologique » (d'où l'expression réalisme ontologique). J'ai choisi de ranger dans cette rubrique tous les réalismes dispositionnels, en dépit du fait qu'on pourrait avoir des raisons de penser que ce ne sont pas des réalismes au sens plein puisque, pour un dispositionnaliste, l'existence des valeurs est intrinsèquement liée aux réactions de l'agent : pour lui, les valeurs ne sont pas parfaitement indépendantes de nos attitudes, croyances, sentiments. Mais il existe plusieurs versions du dispositionnalisme. Certaines admettent

1) Le réalisme moral sémantique montre comment il est possible d'appliquer le prédicat « être vrai » à des énoncés moraux tels que « L'esclavage est un mal », « Il faut tenir ses promesses », exactement de la même façon qu'on l'applique à « $2 + 2 = 4$ », « L'argile est imperméable », « Balzac est né à Tours le 20 mai 1799 ». Cette théorie dit qu'il suffit, pour rendre toutes ces propositions équivalentes, d'adopter une conception non-réductionniste de la vérité, c'est-à-dire une conception qui exclut toute *définition* de la vérité, (en termes de correspondance à la réalité, cohérence, assertabilité justifiée, etc.) et qui ne donne aucun rôle

... une relative autonomie des valeurs à l'égard des réactions émotionnelles, ou autres, de l'agent. D'autres affirment que, du fait que les valeurs ne peuvent être appréhendées que d'une certaine manière ou dans une certaine perspective (émotionnelle, par exemple), il n'en résulte pas qu'elles n'existent pas intrinsèquement, c'est-à-dire même lorsqu'elles ne sont pas appréhendées (cf., par exemple, la défense de Christine Tappolet : « Les émotions et les concepts axiologiques », dans *Raisons Pratiques*, 6 (*La couleur des pensées*), Paris, E.H.E.S.S., 1995. On peut se demander, bien sûr, si le terme dispositionnel convient encore dans ce dernier cas. Apparemment non, mais il me semble qu'il est difficile de comprendre cette thèse si elle n'est pas présentée comme une utilisation ou une extension réaliste de l'argument dispositionnaliste. On peut se faire une bonne idée du problème dans les textes de John McDowell, en particulier dans « Values and Secondary Qualities » dans Sayre-Mc Cord G., éd., *Essays on Moral Realism*, op. cit., 1988, p. 166-180. Quoi qu'il en soit, dans l'ensemble de ceux que j'appelle réalistes ontologiques, certains ne privilégient ni les valeurs, ni l'analogie avec les qualités secondes. C'est plutôt à partir d'une analogie avec les états mentaux qu'ils élaborent leur théorie. Ils ne sont pas exposés aux mêmes objections. Cf. Brink D.O., *Moral Realism and the Foundation of Ethics*, Cambridge, Cambridge University Press, 1989, et la plupart des auteurs cités dans la seconde partie du recueil de Sayre-Mc Cord, op. cit., 1988. Les deux autres réalistes ont des défenseurs attitrés et des appellations spécifiques. C'est Hilary Putnam qui parle de réalisme interne dans *Le réalisme à visage humain* (1990), trad. C. Tiercelin, Paris, Le Seuil, 1994, en développant des arguments défendus, entre autres, dans *Raison, vérité et histoire*, op. cit., 1984, et *The Many Faces of Realism*, La Salle, Illinois, Open Court, 1987. Quant au réalisme normatif, c'est celui que Thomas Nagel propose dans *Le point de vue de nulle part* (1986), trad. S. Kronlund, Combas, L'éclat, 1993.

explicatif à la vérité, (en termes d'utilité ou d'efficacité, etc.) Cette théorie pauvre, austère, minimale de la vérité nous autorise à dire d'une proposition morale telle que « L'expérimentation sur les animaux est un mal » qu'elle est vraie, même s'il ne s'est jamais trouvé personne jusqu'à présent pour la reconnaître comme telle et même s'il ne se trouvera jamais personne à l'avenir pour la reconnaître comme telle. Elle pourrait être vraie, également, même si nous l'avions reconnue comme telle, mais sans avoir aucun moyen effectif de la prouver, ni même une idée de ce que pourrait être une vérification dans ce cas. Bref, cette théorie permet de soutenir que la vérité des propositions morales est parfaitement indépendante de nos capacités de la reconnaître ou de la justifier. C'est en ce sens qu'elle est réaliste, bien qu'elle neutralise les différences entre les domaines de réalité, ou, plus exactement, choisit de ne pas se situer sur le plan ontologique.

2) Le réalisme moral ontologique étend, aux propriétés morales telles que la bonté ou le courage, le statut qu'on attribue ordinairement aux propriétés physiques telles que la rondeur ou la solidité. À première vue, il s'écarte d'une tradition qui fait des propriétés morales des *projections* de nos sentiments et de nos attitudes, au même titre que les propriétés esthétiques (la beauté ne serait pas dans les choses, mais dans nos sentiments à leur égard) ou les qualités secondes (les couleurs ou les saveurs ne seraient pas des propriétés intrinsèques des choses, c'est-à-dire des propriétés que les choses auraient même si nous n'existions pas). Mais en fait, le réaliste moral ontologique se sert de l'argument traditionnel de l'alignement des propriétés morales sur les qualités secondes pour le retourner à son avantage. Il montre qu'il existe des raisons de dire des qualités secondes qu'elles sont des propriétés intrinsèques ou, au moins, dispositionnelles des choses. Et il estime, en conséquence, que rien ne lui interdit d'étendre cet argument aux propriétés morales aussi, puisqu'elles sont supposées avoir le même statut ontologique que les qualités secondes. Il en résulte que, pour lui, dire d'une institution qu'elle est juste ou d'un acte qu'il est bon, ce n'est pas décrire nos attitudes ou nos sentiments à leur égard, mais des propriétés intrinsèques ou, au moins, dispositionnelles de cet acte ou de cette institution, des propriétés qui sont, en un sens important quelconque, indépendantes de ce que nous pensons et ressentons. Il

ne voit pas d'inconvénient, non plus, à attribuer des pouvoirs causaux authentiques aux propriétés morales, dès lors qu'il existe des raisons d'en accorder aux qualités secondes.

3) Le réalisme interne admet, comme le réalisme ontologique, qu'il y a des propriétés normatives ou morales objectives. Mais il estime qu'il y a une grande différence entre dire d'une propriété normative ou morale qu'elle est une propriété *objective* et dire qu'elle est une propriété *intrinsèque* des choses. Or le réaliste ontologique confond les deux. En fait, le réaliste ontologique reste prisonnier d'une dichotomie intenable entre les propriétés intrinsèques des choses et les projections de nos attitudes et de nos sentiments sur ces choses. Pour le réaliste interne, des propriétés normatives telles que la cohérence ou la simplicité sont des propriétés objectives, en ce sens qu'elles sont indépendantes de nos sentiments et de nos attitudes et même de la connaissance que nous pouvons en avoir, mais cela ne veut pas dire qu'elles soient des propriétés intrinsèques de ces choses. On peut étendre cette observation à des propriétés morales telles que la cruauté ou la lâcheté. D'autre part, le réaliste interne estime, comme le réaliste sémantique, que le prédicat « être vrai » peut s'appliquer aussi bien aux énoncés physiques qu'aux énoncés moraux. Mais il ne pense pas qu'il soit indispensable, pour expliquer ce résultat, de vider la notion de vérité de tout contenu (et de la rendre intenable pour cette raison). D'après le réaliste interne, on peut donner un contenu à la vérité qui n'interdit pas son application large. Il pense que tel est bien le cas lorsqu'on définit la vérité en termes d'acceptabilité rationnelle (dans des conditions idéales).

4) Le réalisme normatif affirme que les problèmes de la connaissance empirique et ceux de la connaissance morale ne sont pas incommensurables même s'il existe entre eux d'importantes différences. L'objectivité est, dans les deux cas, le résultat provisoire d'une modification rationnelle de nos intuitions ordinaires, un mouvement de dépassement graduel des perspectives *particulières* vers un point de vue de plus en plus *impersonnel*, impartial, détaché. À l'objection disant qu'il existe dans le domaine moral un conflit souvent dramatique entre les perspectives particulières et les perspectives universelles ou impartiales, le réaliste normatif répond qu'il y a également dans la

connaissance empirique une réelle difficulté à concilier les points de vue subjectif et objectif. Mais l'unification du fait et de la valeur se fait néanmoins au profit du fait, car c'est la tendance naturelle à l'impartialité, propre à notre espèce et magnifiquement illustrée par le développement des sciences empiriques et des mathématiques, qui l'emporte. Et de même qu'il serait absurde de se demander pourquoi nous tendons à l'objectivité dans la connaissance empirique, il serait absurde de se demander pourquoi nous essayons, dans la mesure du possible, d'adopter un point de vue objectif ou impartial lorsque nous agissons. En définitive, le réalisme normatif restaure au plan de la méthode et des idées régulatrices (la vérité, la rationalité) l'unité du domaine physique et du domaine moral, dont il admet qu'elle est peut-être inexistante par ailleurs.

3

On peut avoir des raisons de se demander si le terme réalisme convient bien au réalisme interne. Il offre des perspectives réalistes intéressantes en philosophie morale en affaiblissant considérablement l'idée de projection ou en effaçant la frontière entre propriétés et projections. Cependant, le réaliste interne a plutôt tendance à parler d'acceptabilité rationnelle que de vérité. Ce qui fait qu'on peut lui reprocher de défendre, en définitive, une forme déguisée d'historicisme ou de relativisme (accusation que le réaliste interne a les moyens de réfuter cependant)¹. On peut aussi s'interroger sur la cohérence d'un classement qui range, dans le même ensemble, des formes de réalisme qui portent plutôt sur les propriétés ou les valeurs (les trois premiers) et un autre, qui porte exclusivement sur les normes et les raisons d'agir (le quatrième). Nous verrons qu'il aurait été probablement plus légitime de les classer à part (et peut-être aussi, de ne pas appeler réalistes les théories qui s'occupent uniquement de l'objectivité des raisons d'agir). Mais dans la mesure où aucune de ces formes de réalisme n'est, à proprement parler, un réalisme

1. Cf. les objections de Hartry Field à Hilary Putnam « Realism and Relativism », *The Journal of Philosophy*, 1982, p. 553-567 ; et la réponse de Putnam, à la suite de l'article.

naïf, du genre de celui qui imprègne nos conceptions ordinaires des objets physiques, on peut admettre que le terme soit employé dans un sens assez large.

Les quatre formes de réalisme sont non réductionnistes, en ce sens qu'aucune ne propose une définition restrictive du bien : plaisir, bien-être personnel, utilité sociale, etc. Cependant le réaliste interne a parfois tendance, en raison de sa répugnance à l'égard de toutes les dichotomies, à rejeter les divisions tranchées entre les valeurs morales et les valeurs sociales ou les besoins humains fondamentaux, ce qui peut donner, parfois, un ton réductionniste à ses analyses. Mais, en gros, le programme des réalistes moraux est assez clair : développer de façon cohérente, une théorie non réductionniste de la vérité des jugements moraux. En effet, on pourrait dire qu'un réaliste moral est un philosophe qui est prêt à endosser ces deux thèses au moins¹ :

1) Les énoncés moraux expriment des propositions, qui sont vraies ou fausses objectivement, c'est-à-dire indépendamment de la question de savoir si nous avons reconnu ou pourrions reconnaître leur vérité ou leur fausseté.

2) Les prédicats moraux dénotent des propriétés qui ne sont pas réductibles à d'autres propriétés non morales : « bien » n'est pas équivalent à « plaisant », « utile au plus grand monde », « propre à satisfaire des besoins humains fondamentaux », « propre à satisfaire l'épanouissement personnel », etc.

Si un philosophe admet que des énoncés éthiques peuvent être dits vrais ou faux, sans reconnaître cependant qu'il pourrait y avoir des vérités éthiques qui transcendent nos capacités de les reconnaître, ce n'est pas un réaliste. On dira tantôt que c'est un vérificationniste, tantôt que c'est seulement un cognitiviste (parce qu'il admet que les énoncés moraux peuvent être dits vrais ou faux), anti-réaliste (parce qu'il n'accepte pas qu'il puisse y avoir des vérités éthiques que

1. Ces deux thèses sont minimales en ce sens qu'elles doivent pouvoir être admises par les partisans des quatre variétés distinguées. C'est pourquoi elles ne font pas intervenir de considérations « causales » ou « ontologiques » qui excluraient tous les réalismes à l'exception du réalisme ontologique.

nous n'aurions aucun moyen d'appréhender ou de démontrer). Bref, on ne peut pas être réaliste si on n'accepte pas une version forte de (1)¹.

Si un philosophe estime que certaines formes de réduction ne sont pas incompatibles avec l'autonomie (ou l'irréductibilité ou le caractère *sui generis*) des propriétés morales, il peut, à la rigueur, être dit réaliste. Bref, il semble qu'on puisse être réaliste, même si on n'accepte qu'une version faible de (2). Cependant, le célèbre argument de Moore, dit de la question ouverte, exclut, en un certain sens, cette dernière possibilité².

À première vue, l'argument de Moore est d'une simplicité déroutante. Il dit seulement : toute définition empirique du bien (l'utile, l'agréable, l'impartial, etc.) appelle la question : en quoi cet équivalent empirique du bien est-il un bien ? (En quoi l'utile ou l'agréable ou l'impartial est-il un bien ?) Cette question se pose à nouveau pour toute explicitation de la première définition, d'où le nom d'*argument de la question ouverte*. Il apparaît ensuite que l'argument est un peu plus complexe et un peu moins convaincant.

En fait Moore affirme qu'une phrase telle que « Cette action est bonne, mais elle ne contribue pas au bien-être de la société »,

1. Le débat réalisme/anti-réalisme ne porte pas uniquement sur la question de savoir s'il y a des vérités morales mais aussi sur celle de savoir s'il pourrait y en avoir même si nous ne les avons pas reconnues pour telles ou même si nous n'avions aucune idée de ce qu'on devrait faire pour les prouver ou les démontrer. Le réaliste moral peut fort bien admettre que certaines propositions morales sont susceptibles de démonstration. Mais ce qui caractérise son point de vue, c'est qu'il affirme qu'il pourrait y avoir des vérités non démontrées et non susceptibles de l'être. En revanche, l'anti-réaliste pense qu'il serait absurde de parler de vérités éthiques que nous ne serions pas en mesure de reconnaître ou pour lesquelles nous n'aurions aucune idée du genre de procédure qu'il serait nécessaire d'utiliser pour les appréhender. Un anti-réaliste pourrait donc parler de propositions morales vraies ou fausses, mais en réservant l'usage de ces prédicats aux propositions démontrées ou susceptibles de l'être. C'est pourquoi on peut dire que l'anti-réaliste est un vérificationniste. Mais il est difficile, en philosophie morale, d'être vérificationniste sans être réductionniste, d'où le lien entre les deux thèses (et l'intérêt du réalisme).
2. Moore G.E., *Principia Ethica*, Cambridge University Press, 1903.

n'est pas contradictoire. Il en tire la conclusion qu'il n'y a pas d'identité conceptuelle entre l'idée de bien et celle de bien-être de la société. Mais, pour ses critiques, cette absence d'identité conceptuelle n'exclut pas qu'il puisse exister entre le bien et le bien-être de la société une relation empirique constante, telle qu'on puisse parler d'une identité synthétique ou *a posteriori*, des propriétés « être bon » et « contribuer au bien-être de la société »¹. C'est d'ailleurs un peu celle que Durkheim avait cru établir². Si cette idée est acceptable, une phrase telle que « cette action est bonne mais elle ne contribue pas au bien-être de la société » serait toujours fausse empiriquement, sans être jamais contradictoire. Que le bien ne soit pas le bien-être de la société n'est manifestement pas une impossibilité conceptuelle, mais c'est peut-être une impossibilité empirique, si cette dernière expression a du sens.

Mais, en admettant qu'il existe, en effet, une différence entre l'identité conceptuelle et l'identité des propriétés et que rien, en conséquence, ne nous interdit de penser que nous pourrions trouver empiriquement ou *a posteriori* en quoi consiste cette identité de propriétés, rien ne nous interdit de penser, non plus, que nous ne la trouverons jamais. Et l'une des raisons pour lesquelles il se pourrait que nous ne puissions jamais découvrir cette identité *a posteriori*, c'est que nous ne savons même pas ce que nous cherchons à expliquer ! Pour l'eau et la température, qui sont devenues de véritables emblèmes des définitions physiques ou synthétiques réussies, on sait vaguement ce qu'il y a à expliquer (en gros, des sensations). Mais, pour le bien, le problème n'est pas que nous n'avons qu'une idée vague de ce qui est à expliquer, c'est que nous en avons plusieurs idées incompatibles. Ce n'est pas du tout la même chose d'expliquer ou de définir les sentiments d'approbation ou de désapprobation, les intentions bonnes ou mauvaises, les comportements altruistes ou égoïstes, etc.

En résumé, l'argument de l'identité *a posteriori* des propriétés n'élimine pas la question de Moore, même s'il la déplace un peu. De toute façon, personne n'appellerait « réaliste » un philosophe moral qui afficherait clairement ses préférences pour un réduc-

1. Putnam H., *op. cit.*, 1984, p. 229.

2. Durkheim E., *op. cit.*, 1911.

tionnisme sans nuances (c'est-à-dire, sans certains aménagements du genre de ceux qu'on peut trouver dans la théorie dite de la « survenance » lorsqu'elle est appliquée aux états mentaux. Je reviendrai sur le sujet).

Ces différentes formes de réalisme non réductionnistes peuvent-elles coexister pacifiquement ?

Je ne le crois pas. Le réaliste ontologique rejette la vacuité du réalisme sémantique ; le réaliste normatif rejette les fantaisies métaphysiques du réaliste ontologique ; le réaliste interne rejette en bloc les divisions que présupposent le réalisme ontologique et normatif et la vacuité du réalisme sémantique.

4

La meilleure manière de faire le point sur la question, c'est, me semble-t-il, d'examiner les trois arguments fondamentaux autour desquels la discussion s'est organisée :

- 1) l'argument sémantique austère,
- 2) l'argument de l'analogie entre propriétés éthiques et qualités secondes ou états mentaux ou objets mathématiques,
- 3) l'argument du « holisme » de la connaissance.

1) L'argument sémantique

Dans le cadre d'une sémantique austère, il suffit qu'un énoncé ait une forme descriptive pour qu'on puisse dire de lui qu'il est susceptible de vérité ou de fausseté¹. Or si on peut avoir des doutes à propos des énoncés moraux prescriptifs, ceux qui ont la forme d'impératifs ou ceux qui introduisent des opérateurs déontiques tels que « il faut, tu dois, etc. », ces doutes ne peuvent pas subsister devant les énoncés évaluatifs (ou axiologiques) : « Les notaires sont honnêtes », « Sacrifier des innocents est ignoble » sont des énoncés descriptifs bien formés. On peut tenir pour acquis que le discours moral est, en partie au moins, descriptif. C'est une raison suffisante d'estimer que ce domaine de discours pourrait être dit réaliste, en partie au moins. Pour que le discours moral puisse être dit réaliste, en partie au moins, on devrait être en mesure d'ajouter :

1. Horwich P., *op. cit.*, 1990.

- 1) les énoncés moraux ont une valeur de vérité déterminée
- 2) certains de ces énoncés sont vrais
- 3) le fait qu'il y ait des énoncés moraux vrais est indépendant de nos capacités de le reconnaître, et des moyens effectifs que nous avons de le prouver.

Cette dernière clause (3) pourrait modifier profondément nos vues sur les vérités éthiques. Car, la plupart du temps, c'est l'absence de possibilité de les prouver qui nous fait renoncer à l'idée qu'il existe des vérités de ce genre. C'est cette clause qui nous autorise à dire, par exemple, que la proposition « L'expérimentation sur les animaux est un mal » est vraie même si nous n'avons aucun moyen de le prouver effectivement. Elle pourrait être vraie, d'ailleurs (ainsi que d'autres vérités éthiques auxquelles nous n'avons encore jamais pensé et auxquelles nous ne penserons peut-être jamais !) même si nous restions pour l'éternité incapables de les prouver.

Tout en étant lui-même anti-réaliste, Michaël Dummett a proposé une formulation du réalisme remarquable par sa clarté¹. Certains diront qu'il s'est arrangé pour en proposer une version tellement radicale qu'elle est inoffensive. Mais ce n'est certainement pas l'opinion de la multitude de philosophes qui se sont donnés la peine d'essayer de la repousser.

À la question métaphysique traditionnelle : existe-t-il une réalité indépendante de notre esprit ? Dummett a proposé de substituer la question sémantique suivante : pour une certaine classe d'énoncés, la valeur de vérité de chaque énoncé de la classe est-elle déterminée indépendamment de la connaissance que nous en avons ou pas par quelque réalité objective dont l'existence et la constitution sont, elles aussi, indépendantes de cette connaissance ? De cette façon, Dummett a cru, semble-t-il, qu'il transformait une question confuse et insoluble en une « confrontation entre des positions clairement discernables et rationnellement discutables² ». Être réa-

1. Dummett H., *Truth and Other Enigmas*, Londres, Duckworth, 1978 ; *The Logical Basis of Metaphysics*, Cambridge, Mass., Harvard University Press, 1991 ; *The Interpretation of Frege's Philosophy*, Londres, Duckworth, 1981.

2. Bouveresse J., *Le pays des possibles : Wittgenstein, les mathématiques et le monde réel*, Paris, Minuit, 1988, p. 23.

liste, relativement à une certaine classe d'énoncés, ce serait tout simplement affirmer que ceux-ci ont une valeur de vérité déterminée, que celle-ci ait été appréhendée ou pas, que nous ayons les moyens effectifs de la vérifier ou non. La question ontologique du mode d'existence des objets, de leur dépendance ou de leur indépendance à l'égard de l'esprit, serait secondaire, métaphorique pour ainsi dire. Affirmer que nous « découvrons » une réalité indépendante ou que nous l'« inventons » ne serait qu'une certaine façon de présenter, de façon imagée, notre choix de l'une ou l'autre de ces positions sémantiques.

Ou bien nous estimons qu'il est absurde de dire d'énoncés d'une certaine classe qu'ils ont une valeur de vérité déterminée, lorsque nous n'avons aucune idée de ce qui pourrait compter comme preuve de leur vérité ou de leur fausseté, aucune méthode de vérification disponible. Dans ce cas nous serons, relativement à cette classe d'énoncés, anti-réalistes et nous aurons l'impression que ces réalités sont construites ou inventées, dans le domaine en question, alors qu'en fait, nous serons simplement engagés envers une thèse sémantique anti-réaliste.

Ou bien nous pensons qu'il n'est pas du tout déraisonnable d'affirmer que les énoncés d'une certaine classe ont une valeur de vérité déterminée, même dans le cas où nous n'avons pas (et n'aurons peut-être jamais) de moyens effectifs de les démontrer ou de les vérifier. En d'autres termes, nous estimons que la vérité de ces énoncés peut transcender nos capacités à la reconnaître ou à la démontrer. Dans ce cas, nous serons réalistes relativement à cette classe d'énoncés, et nous aurons l'impression que ces réalités doivent être découvertes (ou plus exactement : la métaphore de la découverte sera plus appropriée que la métaphore de l'invention). La confrontation entre ces deux points de vue, réaliste et anti-réaliste, n'a pas de domaine spécifique. On peut l'envisager aussi bien lorsque nous discutons d'énoncés mathématiques que d'énoncés physiques. Elle peut être déplacée, sans aucune difficulté, vers les énoncés moraux.

Le réaliste moral, au sens dummettien, c'est celui qui pense que l'impossibilité dans laquelle nous nous trouvons de vérifier ou de démontrer nos jugements moraux n'interdit pas qu'ils puissent être réputés vrais ou faux. Et l'anti-réaliste moral, c'est, tout simplement, celui qui nie qu'une telle position soit raisonnable.

Mais, pour de nombreux philosophes, c'est précisément la généralité de la théorie de Dummett qui en est le principal inconvénient. Ils ne sont guère disposés à accepter qu'un argument emprunté à la théorie de la signification puisse nous aider à résoudre la question métaphysique de l'existence des objets¹. D'ailleurs, c'est bien parce que cette théorie admet une sorte d'équivalence entre objets mathématiques et objets physiques qu'elle est suspecte. Il semble aller de soi que les questions d'existence ne coïncident pas avec les questions de vérité. On peut être réaliste « dummettien » (c'est-à-dire du point de vue de la vérité) à la fois pour les objets physiques et les objets mathématiques, tout en étant réaliste de sens commun (c'est-à-dire du point de vue de l'existence) pour les objets physiques seulement. La principale objection à l'argument sémantique c'est donc qu'il n'a aucune implication ontologique et qu'il ne peut, de ce fait, apporter aucune réponse intéressante à la question de la « réalité » des objets d'un domaine particulier.

Quoi qu'on pense de la thèse de la neutralité ontologique de l'argument sémantique en général, il est difficile de nier que cet argument ne peut guère nous éclairer dans le cas qui nous préoccupe et dans d'autres, voisins. Le discours sur les croyances est parfaitement descriptif. On peut envisager la possibilité que chacun des énoncés dans ce domaine ait une valeur de vérité déterminée, même si nous ne la connaissons pas et même si nous ne savons pas comment nous pourrions la connaître. Mais conclure que le discours sur les croyances est réaliste laisse entièrement ouverte la question de la réalité des croyances. Le réalisme sémantique n'est pas incompatible avec le physicalisme. Mais il ne l'est pas non plus avec l'interprétationnisme ou le dualisme, qui nient, en un certain sens, la réalité physique des croyances. Ceux qui estiment que la seule question intéressante à propos des croyances c'est celle de leur réalité physique risquent d'être frustrés par l'argument sémantique. Ils penseront probablement que la neutralité ontologique de l'argument est un inconvénient plutôt qu'un avantage.

1. Devitt M., « Dummett's Anti-realism », *The Journal of Philosophy*, LXXX, 1983, p. 73-99 ; Horwich P., « Three Forms of Realism », *Synthese*, 51, 1982, p. 181-201 ; Bouveresse J., « Frege, Wittgenstein, Dummett et la nouvelle "querelle du réalisme" », *Critique*, 399-400, 1980, p. 881-896.

De la même façon, le discours moral est descriptif, en partie au moins. On peut envisager la possibilité que chacun des énoncés dans ce domaine ait une valeur de vérité déterminée même si nous ne la connaissons pas et même si nous n'avons aucune idée des moyens qui pourraient nous permettre de la connaître. Une interprétation réaliste du discours moral n'est donc pas à exclure. Elle est compatible avec le platonisme moral (il y a des valeurs et des faits moraux non naturels indépendants de nos croyances et de nos préférences) et avec le naturalisme non réductionniste (il y a des valeurs et des faits moraux indépendants de nos croyances et de nos préférences ; ces faits ont une certaine autonomie mais ils sont liés d'une façon quelconque à des faits naturels). Mais elle est compatible, aussi, avec certaines formes de subjectivisme moral qui nient l'existence indépendante des valeurs ou des faits moraux¹. Si la principale ambition du réalisme moral est d'en finir avec le subjectivisme moral, l'argument sémantique ne peut pas la servir.

En résumé, l'argument sémantique peut nous aider à nous convaincre que l'impossibilité dans laquelle nous nous trouvons de vérifier ou de démontrer nos énoncés moraux ne suffit pas à prouver qu'aucun d'entre eux n'est vrai, car les vérités éthiques sont peut-être tout simplement des vérités que nous n'avons pas encore découvertes ou que nous n'avons ni les capacités psychologiques ni les moyens épistémologiques de démontrer. L'argument sémantique devrait nous permettre, entre autres, de contester la *théorie de l'erreur* de Mackie. Mackie reconnaissait que le discours moral était descriptif et qu'il avait, de ce fait, une prétention à la vérité. Mais il en tirait la conclusion (sceptique d'après lui, nihiliste en réalité) que tous les énoncés moraux étaient faux, au même titre que tous les énoncés religieux (d'où le nom de la théorie de l'erreur)². Comment est-il arrivé à ce résultat ? Par vérificationnisme. Mackie juge que tous les énoncés moraux sont faux parce que, d'après lui, nous n'avons aucun moyen effectif de les vérifier au moyen de faits empiriques ordinaires ou

1. Engel P., « La quasi-réalité des valeurs et des normes », *Droits*, 18, 1993, p. 99-107.

2. Mackie J.L., *Ethics. Inventing Right and Wrong*, Penguin Books, 1977.

d'arguments logiques classiques. Mais le réaliste peut objecter que, du fait que nous ne savons pas comment vérifier nos énoncés éthiques ou que nous n'avons pas la possibilité de le faire au moyen des procédures ordinaires, il ne résulte pas que tous les énoncés éthiques soient faux. Il se pourrait que certains de ces énoncés soient vrais bien que nous n'ayons ni les capacités ni les moyens de le démontrer. Il se pourrait aussi que nous n'ayons pas réussi, encore, à formuler un énoncé éthique ce qui ne signifierait nullement qu'il n'y en a pas ou que nous n'en découvrirons jamais. De toute façon, le fait que nous soyons dans l'obligation d'admettre que certaines vérités, au moins, sont saisies directement et ne sont susceptibles ni de démonstration ni de justification empirique (les premiers principes non dérivés) reste un argument assez puissant contre la plupart des formes de vérificationnisme.

De tous ces points de vue, l'argument sémantique est extrêmement fécond (et même assez fascinant, en un certain sens). Mais l'argument sémantique est neutre du point de vue ontologique. Il ne dit rien de la nature de ces réalités auxquelles il se pourrait que nous ne puissions accéder que de façon intuitive (ou auxquelles il se pourrait que nous ne puissions pas accéder du tout !). Il n'est pas incompatible avec un robuste platonisme. Mais il n'est pas incompatible, non plus, avec un certain subjectivisme. C'est surtout cette dernière possibilité qui agace les réalistes moraux. Il n'est pas étonnant qu'ils aient jugé l'argument sémantique plus qu'insuffisant.

2. L'argument de l'analogie

On a vu que les énoncés moraux pouvaient être ou bien évaluatifs ou bien prescriptifs et que les énoncés évaluatifs pouvaient être reconnus à leur forme descriptive et à la présence de prédicats axiologiques généraux tels que « bon », « mauvais », « juste », « injuste » ou spécifiques tels que « courageux », « lâche », « mesquin », « généreux ». Des phrases telles que « Le plombier est honnête », « Mentir est mauvais » sont, selon ces critères, des jugements moraux évaluatifs. Les expressions prescriptives se reconnaissent tantôt au mode (impératif), tantôt à la présence d'indicateurs déontiques tels que « il faut ; il ne faut pas », « Il est permis, obligatoire, interdit », etc. Des phrases telles que « Il faut

tenir ses promesses », « Ne vole pas » sont, selon ces critères, des jugements moraux prescriptifs. Il est possible d'émettre un jugement prescriptif qui ne s'exprime linguistiquement ni à l'impératif ni au moyen d'un indicateur déontique (« Les enfants dorment » pour « ne pas les déranger »). De même, on peut formuler des jugements évaluatifs sans prédicat axiologique (« L'épicier n'a pas rendu la monnaie » pour « L'épicier n'est pas honnête »). Mais le fait qu'on puisse exprimer les évaluations et les prescriptions autrement qu'au moyen des exemples d'énoncés habituellement choisis en philosophie morale ne signifie pas que ces derniers ne nous donnent aucune indication sur la différence entre les jugements évaluatifs et prescriptifs (on peut admettre que ces énoncés indiquent quelque chose à propos de cette différence, même si on défend l'idée qu'en définitive, l'existence de cette différence est indépendante de ses manifestations linguistiques).

C'est, en tout cas, à partir d'une distinction de ce genre que le réalisme dit ontologique semble pouvoir être défendu. Il ne paraît pas absurde de se demander si la propriété axiologique d'être « honnête » est réelle ou aussi réelle que la propriété d'être « blanc ». Mais se demander si la règle disant que « il ne faut pas mentir » est aussi réelle que la propriété de « être une ligne droite » pourrait ressembler à un égarement momentané de l'esprit.

Presque personne d'ailleurs ne semble penser qu'un réalisme de type ontologique pourrait être défendu de façon aussi convaincante pour les normes et pour les valeurs. Et si certains philosophes ont choisi d'appeler leur réalisme « normatif », c'est probablement parce qu'ils s'intéressent plus aux normes et aux raisons d'agir qu'aux prédicats axiologiques, et qu'ils estiment qu'il est absurde d'utiliser pour défendre un réalisme des normes, des arguments du type de ceux que les réalistes axiologiques emploient.

D'une certaine façon, la justification du réalisme moral ontologique au moyen de la division entre valeurs et normes l'expose à une objection dévastatrice. Il suffirait de montrer que les valeurs morales sont liées de façon non contingente aux normes pour ruiner la possibilité d'un réalisme de valeurs. Car si les normes ne sont pas susceptibles d'être traitées de façon réaliste au sens ontologique et si les valeurs ne sont que des sortes de normes, on ne voit pas très bien ce qui pourrait justifier un réalisme ontologique des valeurs.

Mais on ne peut pas exclure l'inverse non plus, c'est-à-dire la dépendance des normes à l'égard des valeurs et la possibilité de justifier un réalisme ontologique des règles à partir d'un réalisme ontologique des valeurs. C'est à peu près de cette façon que le réaliste moral ontologique écarte cette objection préjudicielle. Il lui reste à montrer que l'analogie entre les qualités secondes et les propriétés morales est légitime et qu'elle suffit à justifier la réalité objective des propriétés éthiques.

Le réaliste moral ontologique a une évidente prédilection pour la comparaison entre le fait de percevoir que quelque chose est bien ou mal et le fait de percevoir que quelque chose est blanc ou rouge, bénéfique ou dangereux, tragique ou comique¹. C'est une analogie inspirée de Hume vraisemblablement. Le réaliste moral estime probablement qu'il peut exploiter l'argument de Hume à son avantage.

Dans la tradition de Hume, on estime que les qualités ou les propriétés morales ont la même nature ontologique que les qualités secondes (couleurs, sons, saveurs, etc.). Elles ne sont pas dans les choses, elles sont en nous. Elles ne sont pas à découvrir, elles sont projetées. Ce sont des sortes d'illusions identiques à celles qu'on a lorsqu'on pense que la beauté d'un objet est en lui et non dans l'intérêt ou l'amour que nous lui portons. Mais admettons qu'on propose des arguments qui montreraient que les couleurs, les sons, les saveurs sont, en un certain sens (non négligeable), dans les choses, à titre de dispositions. On justifierait l'idée que les couleurs, par exemple, sont des propriétés qui ne pourraient apparaître qu'à certains êtres pourvus d'un équipement de détection approprié (d'une réceptivité adéquate). Ce serait une façon de dire que ces qualités sont réelles, qu'elles sont dans les choses autant qu'elles sont en nous. Si l'argument est correct, tous ceux qui se sont engagés en admettant que les valeurs morales sont identiques aux qualités secondes seront dans l'obligation d'admettre qu'elles sont dans les choses aussi.

Telle est, en gros, la stratégie argumentative du réaliste, celle qui lui permet de retourner contre l'anti-réaliste, l'argument de

1. Mc Dowell J., « Values and Secondary Qualities », dans Sayre-Mc Cord G., *Essays on Moral Realism*, *op. cit.*, 1988, p. 166-181 ; Pettit P., « Realism and Response-Dependence », *Mind*, 4, 1991, p. 586-626.

l'analogie. Mais le réaliste ne justifie à aucun moment, de façon indépendante, le choix de l'analogie. Il ne nous donne aucune raison de préférer cette analogie à celle qui mettrait en relation le fait de percevoir que quelque chose est bien ou mal et le fait de percevoir que quelque chose est sacré par exemple.

Pourtant le choix de l'analogie détermine en grande partie les conclusions ontologiques. S'il existe quelques raisons de penser qu'un pont pourrait être dangereux même si personne n'avait conscience du fait qu'il est dangereux, il est beaucoup plus difficile d'envisager la possibilité qu'une vache pourrait être sacrée même si personne n'avait conscience du fait qu'elle est sacrée. En d'autres termes, une analyse purement externaliste de la propriété d'être sacré n'aurait aucun sens. Une analyse externaliste met en relation certaines propriétés physiques de l'objet avec certaines propriétés physiques de l'agent, sans tenir compte de la conscience que l'agent a ou pourrait avoir de cette relation. Rien nous interdit de proposer une analyse externaliste de la propriété d'être dangereux. Mais personne ne prendrait le risque de proposer une analyse purement externaliste de la propriété d'être sacré.

Tous ceux qui pensent que la propriété la plus proche des propriétés éthiques est celle d'être sacré (ils sont nombreux et leur point de vue est loin d'être déraisonnable) auront du mal à s'intéresser aux arguments du réaliste moral ontologique, s'il n'a rien d'autre à offrir qu'une analogie avec les propriétés de couleurs ou celles d'être dangereux ou comique.

De toute façon, ce genre de réalisme ne permet pas de concevoir l'existence de réalités parfaitement indépendantes des croyances et des attitudes de l'agent moral. Il permet, au plus, de soutenir qu'il existe des réalités dépendantes des réactions de l'agent.

Mais il existe une forme de réalisme ontologique un peu plus forte qui s'appuie plutôt sur une comparaison entre propriétés morales et états mentaux ou plus exactement qui cherche à exploiter les arguments en faveur de la réalité des états mentaux pour justifier la réalité des propriétés morales.

Dans l'ensemble foisonnant de ces arguments, certains semblent être en mesure de justifier un réalisme moral ontologique plus fort que le réalisme des qualités secondes, qui reste un réalisme dépendant des « réactions » de l'agent (l'objectivité des pro-

priétés morales reste toujours plus ou moins fonction de certaines facultés de réception de l'agent). Pour ce qui concerne les états mentaux, on peut admettre plus aisément qu'ils existent ou n'existent pas indépendamment de nos croyances ou de nos attitudes : c'est particulièrement évident dans le cas des croyances des animaux, dont l'existence ou l'inexistence semblent indépendantes des idées que nous nous faisons à ce sujet.

Bien entendu, un interprétationniste en matière d'états mentaux niera que ces derniers aient une existence indépendante. D'autre part, certains réalistes estiment que la croyance en l'existence ou l'inexistence des croyances d'autrui implique au moins la croyance en l'existence de ses propres croyances. De sorte, que, en un certain sens, on peut dire aussi des états mentaux qu'ils sont dépendants de nos réactions. Mais ce type de dépendance épistémologique n'a pas grand-chose à voir avec la dépendance aux réactions de l'agent de l'existence des couleurs ou d'autres qualités secondes. La dépendance épistémologique est ou bien beaucoup plus forte, puisqu'elle peut suffire à nier l'existence des états mentaux (mais dans ce cas, on abandonne complètement l'option réaliste) ou bien beaucoup plus faible (le fait que pour envisager l'existence indépendante des croyances d'autrui on ait besoin d'affirmer l'existence de nos propres croyances ne dit pas que les croyances d'autrui n'existent pas indépendamment des nôtres).

Soit, donc, l'analogie états mentaux-propriétés morales. L'anti-réaliste observe que les propriétés morales sont souvent évoquées dans des explications d'allure causale (« Les fonctionnaires se sont révoltés parce que l'impôt était injuste »), ce qui tend à accréditer l'idée que ces propriétés sont réelles (en vertu de ces pouvoirs causaux). Mais il estime qu'il s'agit d'une illusion que les philosophes de l'esprit connaissent bien¹. Les seules explications causales légitimes sont physicalistes et celui qui veut attribuer des pouvoirs causaux aux états mentaux est confronté à un

1. Blackburn S., « Supervenience Revisited » dans Hacking I., *Exercises in Analysis, Essays by Students of Casimir Levy*, Cambridge, Cambridge University Press, 1985, p. 47-67 ; Harman G., « Moral Explanations of Natural Facts — Can Moral Claims be Tested Against Moral Reality ? », *The Southern Journal of Philosophy*, Supplément, 1986, p. 57-78.

dilemme. Ou bien, il réduit les états mentaux à des états physiques (des états cérébraux, par exemple) : dans ce cas, il sauve le pouvoir causal des états mentaux, mais il élimine leurs caractères spécifiques. Ou bien, il sauve la spécificité de l'état mental, mais il doit renoncer à lui donner un pouvoir causal. D'après l'anti-réaliste, si un philosophe moral veut aligner les propriétés morales sur les propriétés mentales, il se heurtera forcément à la même difficulté.

Le réaliste, de son côté, accepte l'analogie, en pensant qu'elle peut tourner à son avantage, s'il emprunte à la philosophie de l'esprit certaines notions plus ou moins fonctionnalistes¹. Le réaliste reconnaît qu'un état mental ne peut être dit réel que s'il a des pouvoirs causaux et que les seuls pouvoirs causaux concevables sont physiques. Mais il estime qu'il n'est pas impossible, à partir de ces prémisses, de sauver à la fois la spécificité des états mentaux et leurs pouvoirs causaux, c'est-à-dire leur réalité. Tout d'abord, il affirme que les états mentaux se superposent, pour ainsi dire, à des états physiques (des états du cerveau par exemple) : les propriétés physiques ont une priorité ontologique, dans le sens où elles existent indépendamment des autres propriétés non physiques, alors que les propriétés non physiques n'ont pas d'existence indépendante. La relation des états mentaux aux états physiques est une relation de dépendance asymétrique ou de *survenance*². Ce qui est à la base, c'est l'état physique (à deux états physiques parfaitement indiscernables ne peuvent se superposer deux états mentaux distincts). Mais ce genre de relation de dépendance asymétrique n'implique nullement qu'il y ait *identité* entre la propriété de base (physique) et la propriété superposée (mentale). Les fonctionnalistes nous ont habitués à l'idée qu'une même propriété (mentale par exemple) pouvait être « réalisée » ou « instanciée » dans une disjonction d'états physiques (cet état cérébral, ou celui-là, ou celui-là, etc.). On pourrait dire que le fonctionnaliste pense à l'état mental comme à un « rôle », défini en terme de « besoin » ou de « désir »

1. Brink D.O., *op. cit.*, 1989.

2. Blackburn S., *op. cit.*, 1985 ; Audi R., « Mental Causation : Sustaining and Dynamic » dans Heil J. et Mele A., eds, *Mental Causation*, Oxford, Clarendon Press, 1993, p. 53-74.

de l'organisme, un rôle qui pourrait être tenu (ou incarné) par des « individus » concrets différents. Cet argument permet de sauver la spécificité de l'état mental (puisque'il n'y a pas d'identité entre les types d'états mentaux et les types d'états physiques), ainsi que ses pouvoirs causaux (puisque à tout état mental particulier correspond un état physique particulier). Or sauver la spécificité et les pouvoirs causaux d'une propriété, c'est, en un certain sens, affirmer sa réalité. Le réaliste n'a plus qu'à conclure *ad hominem*, en prenant l'anti-réaliste à son propre piège : les propriétés morales étant comme des propriétés mentales, elles sont aussi réelles que ces dernières, c'est-à-dire irréductibles à des propriétés non-morales, mais superposées ou asymétriquement dépendantes de ces propriétés et, de ce fait, douées d'efficacité causale.

C'est donc au nom de l'équité argumentative que le réaliste moral prétend qu'il est légitime d'attribuer des pouvoirs causaux aux faits moraux ou aux valeurs morales, dès lors que de tels pouvoirs sont accordés aux états mentaux, dont le statut ontologique n'est pas moins incertain. Certains réalistes moraux estiment d'ailleurs que les hiérarchies ontologiques perdent beaucoup de leur pertinence dans l'explication causale. Dire d'un certain genre de faits qu'ils ont une valeur causale, ce n'est rien d'autre, après tout, que dire de ces faits qu'ils donnent, dans un certain contexte, la *meilleure explication possible* des phénomènes examinés. Le fait « moral » qu'une certaine institution est injuste peut être la meilleure explication possible de la réprobation qu'elle suscite¹. Elle est tout à fait satisfaisante en tant que telle. Des principes très généraux d'économie heuristique excluent la nécessité de réduire le fait moral explicatif (l'injustice de l'institution) à d'autres faits non moraux, dont la valeur causale serait, à première vue, moins problématique, mais dont il pourrait apparaître, ensuite, qu'ils le sont tout autant, ce qui entraînerait l'obligation de nouvelles réductions, en nombre indéfini (pensons à la réduction du « fait » de l'injustice de l'institution aux « sentiments » d'injustice des individus qui la réprouvent. Ne prépare-t-elle pas une nouvelle réduction, celle des « sentiments » d'injustice ?).

1. Quinn W.S., « Truth and Explanation in Ethics », *Ethics*, 96, 1986, p. 524-544.

Au fond, le réaliste moral n'est peut-être pas si exigeant. Il observe que, pour les qualités secondes, les états mentaux, les objets mathématiques, on admet qu'il est loisible d'opposer des points de vue non-réductionniste, dispositionnel ou platoniste à des points de vue réductionniste, projectiviste, constructiviste. Et il demande : pourquoi n'en irait-il pas de même pour les valeurs morales ? Pourquoi seuls les points de vue projectiviste ou réductionniste seraient-ils légitimes en ce domaine alors qu'ils ne sont pas les seuls à l'être ailleurs ? Et comme la tendance actuelle est plutôt au réalisme pour les qualités secondes ou les états mentaux, il en tire parti pour revendiquer l'extension du réalisme aux propriétés morales aussi¹.

Toutefois, il ne faudrait pas exagérer l'unanimité des réalistes sur ces sujets. Ainsi la comparaison entre valeurs morales et objets mathématiques n'a pas simplement pour fonction de renforcer le réalisme des valeurs en général, mais de ruiner la comparaison entre qualités secondes et valeurs morales. Opter pour l'analogie avec le domaine mathématique c'est rejeter l'idée impliquée dans l'analogie avec les qualités secondes, que le jugement moral juste ou correct ne serait rien d'autre que la réaction « normale » à un certain genre de stimulation de la part d'un certain type de sujet. C'est essayer de restituer, dans l'analogie, le caractère complexe, réfléchi, de nos évaluations morales².

On retrouve peut-être à cet endroit la difficulté latente du réalisme ontologique. Le réaliste ontologique a tout intérêt à établir ses analogies à partir des énoncés moraux axiologiques. Aussitôt qu'il essaie de rendre compte des énoncés de type normatif (ou prescriptif) il est en difficulté. C'est peut-être parce que les objets

1. Cette tendance assez profonde au réalisme des états mentaux a deux sources apparemment contradictoires. Elle semble refléter, d'une part, l'abandon du behaviorisme classique et, d'autre part, le discrédit qui frappe désormais les dualismes analytique-synthétique, *a priori-a posteriori*, questions de mots-questions de fait, etc. C'est-à-dire, d'une part, une critique de Quine et, d'autre part, une adhésion aux thèses de Quine. Expliquer comment ces deux sources convergent (ou ne convergent pas véritablement) dans la philosophie de l'esprit présente m'entraînerait beaucoup trop loin (ou m'amènerait à me perdre complètement !); c'est un sujet que je n'aborderai pas ici, par conséquent.

2. Virvidakis S., *op. cit.*, 1989.

mathématiques ressemblent plutôt à des règles ou à des normes que le réaliste moral ontologique préfère ne pas s'engager dans la comparaison. C'est d'ailleurs plutôt le réaliste normatif qui s'en occupe¹.

Il n'est pas impossible de tirer toute l'analyse objectiviste des valeurs du côté des normes ou des prescriptions ou plus généralement des raisons d'agir. Le réaliste normatif dira que les chances de donner un sens à l'idée de vérités éthiques sont nulles si on s'en tient à des analogies aussi pauvres et inappropriées que celles qui mettent en relation le sentiment de l'obligation ou de l'interdiction et la perception directe d'une forme ou d'une couleur. L'éthique nous parle seulement de ce qui doit être ou ne pas être, c'est-à-dire précisément de ce qui n'est pas sous nos yeux ou dans nos sens.

Les seules questions intéressantes qu'on peut se poser à propos de l'éthique portent sur nos raisons d'agir. Et le seul réalisme sérieux en matière morale serait celui qui pourrait justifier l'idée qu'il existe des raisons d'agir impartiales, valables pour tous et chacun, indépendamment de ses désirs ou de ses intérêts personnels ou immédiats, ou la thèse disant que les énoncés prescriptifs tels que « Il faut tenir ses promesses », « Il ne faut pas sacrifier des innocents » peuvent être dits vrais ou faux plutôt que rationnellement acceptables ou inacceptables ou acceptables ou inacceptables dans un certain contexte.

Ce genre de théorie pose de formidables problèmes, entre autres : des raisons d'agir impartiales semblent dépourvues d'une propriété importante que devrait avoir, en principe, tout jugement moral, la capacité de motiver ou d'entraîner l'action. Il y a deux moyens de contourner la difficulté : ou bien affirmer que le jugement moral ne doit pas nécessairement motiver ou entraîner l'action, ou bien montrer que des raisons impersonnelles ou des croyances objectives peuvent, à elles seules, sans l'intervention de désirs particuliers, motiver ou entraîner l'action. Ces deux arguments sont directement dirigés contre les thèses de David Hume, mais Hume continue d'avoir de sérieux défenseurs !

D'autre part, il n'est pas du tout sûr que le réalisme des normes puisse être autre chose qu'un intuitionnisme moral. Comment en effet justifier, autrement qu'en faisant appel à une mystérieuse

1. Nagel T., *op. cit.*, 1993.

faculté d'intuition, l'idée que les premiers principes moraux seraient vrais ou faux plutôt que rationnellement acceptables ou inacceptables, ou acceptables ou inacceptables dans un certain contexte ?

Ce sont les premiers principes qui sont en cause car on peut admettre, à la rigueur, que certaines normes puissent être réputées vraies ou fausses selon qu'elles sont ou non dérivées de ces principes. Mais, à moins de se satisfaire d'une forme de cohérentisme, inacceptable en réalité (car il se pourrait que tous les énoncés éthiques soient cohérents entre eux et faux dans l'ensemble), il faut bien justifier ces principes ultimes. C'est à cette justification que sert l'idée d'intuition morale.

Mais le problème des intuitions de premiers principes moraux a toujours été qu'elles ne sont pas unanimement partagées (contrairement aux « intuitions mathématiques »). La question est de savoir, bien entendu, si le fait que ces intuitions morales ne sont pas unanimement partagées n'est pas servi comme un mauvais argument empirique contre l'intuitionnisme. L'intuitionniste pourrait aussi objecter qu'il est faux de dire que ces intuitions ne sont pas partagées. Seuls les relativistes naïfs soutiennent encore qu'il n'y a aucune unité des intuitions morales fondamentales. Mais, à supposer que cette unité de l'appréhension des premiers principes existe (même si elle est masquée par la diversité morale empirique), il n'est pas du tout sûr que la meilleure explication qu'on puisse donner de cette unité soit l'existence d'une faculté d'intuition morale. Il existe beaucoup d'autres arguments qui expliquent correctement cette unité sans faire appel à l'existence d'une faculté spéciale d'intuition (je pense, en particulier, à l'idée de « forme de vie », qui n'est peut-être pas très claire, mais qui a l'avantage de ne pas être aussi « mentaliste » que l'idée d'intuition morale).

D'autre part, les doctrines holistes semblent être, elles aussi, en mesure d'expliquer cette unité, sans faire intervenir l'idée d'une faculté spéciale d'intuition morale. D'après elles, nos facultés ordinaires suffisent, car il n'existerait en réalité aucune différence de nature entre la réflexion morale et la recherche scientifique. Telle est du moins l'interprétation courante de l'argument holiste, que son inspirateur, Quine, est loin de partager cependant.

3. L'argument holiste

Ce sont principalement les réalistes internes qui ont tendance à invoquer la doctrine dite holiste de Quine pour justifier leurs arguments. Mais, en un certain sens, tous les réalistes semblent avoir une dette à l'égard du holisme. Car c'est la doctrine holiste qui a le plus puissamment contribué à effacer les frontières entre le fait et la valeur, au bénéfice du fait. Telle est du moins l'opinion générale des philosophes qui sont souvent très étonnés d'apprendre qu'en réalité Quine, lui-même, refusait cette extension de sa doctrine.

Le holiste dit qu'il est absurde d'essayer d'établir une frontière nette entre nos croyances factuelles et nos croyances non factuelles (théoriques, normatives, etc.). C'est une conséquence de sa critique des « dogmes de l'empirisme », c'est-à-dire de la dichotomie entre l'analytique et le synthétique et de la thèse selon laquelle on pourrait « vérifier » ou « réfuter » un énoncé isolé ou indépendamment de tous les autres auxquels il est lié. C'est aussi une conséquence de sa critique de la théorie positiviste du « double standard ». Pour le positiviste logique, l'ontologie est le choix non vérifiable d'un cadre conceptuel. Seules les hypothèses scientifiques seraient sujettes à vérification. Pour le holiste, la science et l'ontologie ne peuvent pas être distinguées aussi radicalement. L'ontologie n'est pas moins sujette à révision que la science, même si elle occupe une position un peu plus centrale. L'ontologie est indirectement testée par les succès des hypothèses scientifiques qui en sont dérivées.

On peut avoir l'impression qu'il y a une contradiction dans la doctrine holiste entre l'idée que certaines croyances sont tellement centrales qu'on préférera toujours les « sauver », parfois au détriment de faits et de données et l'idée que tout, dans l'ensemble que forment nos croyances, est susceptible d'être révisé, les croyances centrales aussi bien que les autres. Mais centralité des croyances ne veut pas dire vérités éternelles. C'est tout l'intérêt du holisme à la Quine (et l'expression du pragmatisme qui l'anime) de parvenir à expliquer la différence entre nos croyances logiques et les autres, sans traiter les premières comme des vérités immuables.

Toutes ces considérations devraient, en principe, s'appliquer au domaine éthique. Il n'est pas difficile de comprendre pourquoi

elles devraient avoir sur la philosophie morale des positivistes des effets aussi dévastateurs que sur leur philosophie de la science.

Les positivistes logiques avaient essayé de donner une justification définitive à l'un des « dogmes » des sciences humaines : la dichotomie du fait et de la valeur. Ou, plus exactement, à l'une des versions de cette dichotomie, celle qui exclut toute argumentation rationnelle sur les fins ultimes (sur ce qu'on doit faire, le genre de vie qu'on veut mener, le genre de personne qu'on veut être) et toute possibilité d'évaluer en termes de vérité ou de fausseté des énoncés moraux tels que « L'esclavage est un mal ».

Mais la position des positivistes perdait sa solidité à partir du moment où il était démontré qu'il n'y avait aucun sens à discuter de la vérité ou de la fausseté d'un *énoncé isolé*. Un holiste devrait affirmer, en principe, que la vérité ou la fausseté d'un énoncé tel que « L'esclavage est un mal » ne peut pas être déterminée indépendamment de l'évaluation de la vérité ou de la fausseté d'une infinité d'autres croyances dont la « factualité » est indéniable (celles qui portent sur les besoins physiologiques, en particulier). D'autre part, on peut envisager la possibilité que nos croyances éthiques occupent une position similaire à celle des croyances logiques et c'est pourquoi le problème de leur vérification ou de leur réfutation serait si complexe (on préférera, parfois, renoncer à certains « faits » pour sauver nos croyances morales). Mais, de même que nos croyances logiques sont difficilement discutables sans être à l'abri d'une révision imposée par des découvertes non logiques (physiques par exemple), de même nos croyances morales sont protégées sans être immunisées à l'égard de changements que pourraient imposer des découvertes non morales (sociales, physiques, etc.). De tout ceci, il devrait résulter, en principe, que le holisme exclut toute division tranchée entre le fait et la valeur, la science et l'éthique.

Mais Quine n'a jamais dit que son holisme épistémologique, dont l'objectif était plutôt d'intégrer les disciplines dites analytiques (telles que la logique) et les disciplines dites synthétiques (telles que la physique) dans une même entreprise d'accroissement de nos connaissances concernait de quelque façon que ce soit les croyances morales, qu'il juge dépourvues de toute « base observationnelle », fût-elle indirecte. Pour Quine, comparée à la

science, l'éthique est « méthodologiquement infirme » et, comparées à celles de la science, ses prétentions à l'objectivité sont déplacées¹.

Ce n'est pas sans une certaine ironie, d'ailleurs, qu'il remarque que, sur un sujet aussi vital que la différence entre le bien et le mal, on ne dispose absolument pas des instruments de contrôle, dont on se sert massivement pour tester certaines théories scientifiques dont l'intérêt est loin d'être aussi évident.

Bref, Quine semble estimer que les arguments holistes, même pris en totalité — impossibilité de vérifier ou de réfuter un énoncé isolé, enchevêtrement des croyances factuelles et non factuelles, continuité entre les croyances centrales et les croyances périphériques, etc. — ne suffisent pas à justifier la dissolution de la frontière science-éthique.

À première vue, l'attitude de Quine est étrange. Le holisme semble impliquer que c'est l'ensemble de nos croyances qui est engagé chaque fois que nous voulons tester l'une d'entre elles. C'est un argument puissant contre la dichotomie du fait et de la valeur puisqu'il suggère que nous pouvons, indirectement au moins, mettre nos croyances éthiques à l'épreuve. Ce n'est peut-être pas le meilleur argument en faveur du réalisme moral strictement « non réductionniste », mais c'est un argument suffisant, pour des formes de réalisme moral qui admettent un certain degré de réduction (le réalisme interne, par exemple, mais il y en a d'autres que je n'ai pas examinées ici).

Mais Quine récuse entièrement l'idée que les croyances éthiques pourraient être, fût-ce indirectement, « testées ». Pourquoi ? Il semble que Quine refuse d'aligner le discours éthique sur le discours scientifique ou d'intégrer l'éthique dans l'entreprise d'accroissement de nos connaissances parce que le discours éthique n'a pas du tout la même structure d'ensemble que le discours scientifique. Dans le discours scientifique, il y a bien, au centre, des énoncés qui ne sont pas directement confrontés à l'expérience. Mais la ligne qui va de ces énoncés aux énoncés

1. Quine V.V.O, « On The Nature of Moral Values », *Theories and Things*, Cambridge, Mass, Harvard University Press, 1981, p. 56-66. La suite de mon examen repose sur ce texte très bref. Il faut dire que Quine n'a guère écrit sur l'éthique, pour des raisons que ce texte justifie assez bien.

observationnels est continue. L'enchevêtrement est tel qu'on peut dire que les énoncés mathématiques eux-mêmes sont indirectement testés (par les succès de la physique par exemple). À la périphérie du discours scientifique, on trouvera toujours des énoncés observationnels. Mais le discours éthique ne contient et ne peut contenir aucun énoncé observationnel. Toutes les « vérités » qu'on pourrait y trouver, éventuellement, seraient de l'ordre de la cohérence. Mais un holiste n'admet pas de vérité de ce type : en dépit de ce qu'on dit parfois, holisme ne signifie pas cohérentisme.

Quine propose trois arguments qui s'emboîtent pour justifier son point de vue. Le premier porte sur les procédures qui sont utilisées dans la science et dans l'éthique ; le second, sur le genre de fait qu'on recueille dans les deux domaines ; le troisième, sur la genèse des concepts cognitifs et moraux.

D'après Quine, le raisonnement moral n'a rien à voir avec le raisonnement scientifique, car il ne procède jamais par hypothèses et vérifications. Lorsque deux scientifiques s'opposent sur un point particulier, ils acceptent néanmoins de se situer, d'une façon ou d'une autre, par rapport à une réalité indépendante à laquelle ils font appel pour les départager.

Mais il n'en va pas de même en cas de désaccord moral. On peut être en désaccord moral ou bien parce qu'on s'oppose sur les principes (faut-il dire la vérité en toutes circonstances ?) ou bien parce qu'on s'oppose sur le point de savoir si telle ou telle action est une instance du principe (a-t-il *vraiment* menti ?). Dans les deux cas, nous faisons appel, dans nos raisonnements, à une certaine idée de la vérité-cohérence. Dans aucun des deux cas, nous ne testons des hypothèses relativement à une réalité indépendante. Lorsque nous essayons de savoir si tel ou tel principe est moral, nous ne nous posons que des questions de cohérence (le principe est-il compatible avec d'autres principes moraux que nous avons accepté ?). Lorsque nous voulons savoir si nous sommes confrontés à un cas de mensonge, nous essayons de classer un acte particulier selon certains critères. Le travail de classification n'a rien en commun avec la procédure de vérification d'une hypothèse ou d'une prédiction.

Mais ces arguments ne suffisent pas encore à prouver qu'il est exclu qu'il y ait des énoncés observationnels à la périphérie du

discours éthique. Quine ajoute donc un argument spécialement destiné à le montrer. Il pose la question suivante : l'énoncé moral occasionnel « C'est scandaleux », prononcé par une personne qui voit quelqu'un frapper un infirme, peut-il être considéré comme un énoncé d'observation ?

Quine semble penser que dans le cas d'un énoncé observationnel tel que « Il pleut » (lorsqu'il pleut) le locuteur n'a pas, dans des circonstances normales, à justifier son affirmation, si l'auditeur est témoin de la scène (l'auditeur pourrait même reprocher au locuteur de parler pour ne rien dire). Dans le cas d'un énoncé occasionnel tel que « C'est scandaleux », les choses se présentent de façon différente. Il ne paraît pas absurde de demander des justifications ou des explications. Et, au cas où elles sont exigées, on ne peut pas se contenter de répondre « Tu le sais aussi bien que moi, puisque nous observons la même scène » (ou du moins, ce genre de réponse exigerait qu'on ait des convictions réalistes extrêmement fortes au sujet des valeurs, des convictions que Quine trouverait probablement absurdes).

Pendant, Quine conclut que retenir le critère de l'« information collatérale » comme il le fait, c'est-à-dire l'information qui n'est pas donnée dans la scène, c'est faire du statut observationnel des énoncés occasionnels une affaire de degré. On pourrait, d'après lui, ordonner l'ensemble des énoncés occasionnels selon leur degré d'observationnalité (si on peut dire). À l'un des extrêmes, on trouverait des énoncés tels que « C'est un célibataire », qui exigent une information collatérale considérable. À l'autre extrême, il y aurait des énoncés tels que « Il pleut » qui n'en requièrent aucune. « C'est scandaleux » se situerait entre les deux.

Tout cela suffit-il à justifier l'idée que l'éthique et la science ne sont pas dans le même bateau ? Probablement pas puisque Quine ne s'en est pas contenté. En fait, il a cru nécessaire d'apporter une explication beaucoup plus générale (et beaucoup plus contestable) à son refus de faire marcher ensemble l'éthique et la science.

D'après Quine, il existerait une structure minimale de préférences, laquelle se manifesterait dans le fait que, dès l'origine, nous cherchons certaines choses et nous en fuyons d'autres. Ces comportements seraient comme l'expression d'instincts ayant

une valeur de survie (désirer goûter le lait maternel est un bon exemple de ces préférences minimales). Mais les choses ne restent pas dans cet état rustique et de toute façon ces préférences n'ont rien de « moral ». Des préférences dites « morales » semblent se dessiner aussitôt que nos moyens de satisfaire certains besoins ou certains désirs se transfigurent en fins. On pêche d'abord pour attraper des poissons et les manger, c'est-à-dire pour des raisons instrumentales. Mais il peut se faire qu'on s'intéresse graduellement à l'acte de pêcher en le détachant de sa finalité instrumentale initiale. Les moyens deviennent des fins. Nous pêchons pour le plaisir de pêcher, si on peut dire. Le plaisir de la pêche devient désintéressé. On est sur la voie de la moralité. Plus exactement, on adopte, dans certaines activités, une sorte d'attitude morale. Mais ce qui est authentiquement moral a un domaine spécifique : celui des relations avec autrui. C'est dans ce domaine qu'il faut adopter des attitudes désintéressées.

À première vue, rien ne nous interdit d'envisager la possibilité que des comportements inspirés de ces attitudes aient une sorte de base instinctive (la capacité d'éprouver de la sympathie) et certaines valeurs de survie (les plus coopérants l'emportent !).

Mais il est plus difficile de trouver une base instinctive ou une valeur de survie aux principes que nous trouvons éminemment moraux, le meilleur exemple de ces principes étant celui qui nous « oblige » à dire la vérité, alors que la ruse et les pieux mensonges ont des valeurs de survie évidentes.

Quine pense qu'il faut utiliser deux théories différentes pour expliquer cette disparité entre comportements de base et comportements moraux. Pour les comportements de base, les théories du *développement cognitif* sont assez appropriées ; elles donnent aux « instincts » et aux processus « naturels » la place qui convient. Pour les comportements moraux, en revanche, il faut appliquer assez étroitement une *théorie de l'apprentissage*, puisque tout semble acquis par récompenses et punitions.

C'est une discontinuité théorique qui n'a pas d'équivalent lorsqu'on s'intéresse à la genèse des concepts scientifiques. Il n'y a pas de discontinuité radicale entre le processus naturel du développement de la perception des objets et de relations entre objets et la recherche scientifique. La théorie du développement cognitif suffit à rendre compte de tout ce processus.

On pourrait dire, pour être un peu plus précis, que, d'après Quine, on a besoin de deux théories complètement différentes pour expliquer les valeurs morales. L'une pour la genèse et l'autre pour la stabilisation, alors qu'une théorie suffit pour expliquer nos croyances scientifiques, car celles-ci ne sont pas en rupture totale avec nos croyances élémentaires.

La question de savoir si les théories que Quine propose sont les bonnes reste ouverte. On peut admettre avec Quine que ce n'est pas la même chose d'expliquer les valeurs de survie élémentaires et les valeurs morales, tout en rejetant l'idée que la meilleure explication de l'existence de valeurs morales, c'est le behaviorisme, au sens de théorie de l'apprentissage par récompenses et punitions.

À la manière de Hume, Quine estime qu'il y a discontinuité entre la structure minimale des préférences individuelles, les sentiments de sympathie et les institutions morales qui s'expriment dans les sanctions. Mais il est difficile de nier qu'il existe une sorte d'uniformité dans les valeurs morales ou dans les fins ultimes. Si tout se fait par apprentissage et sans aucune relation avec un développement cognitif naturel-universel, comment expliquer cette uniformité des valeurs morales les plus générales ou des fins ultimes ?

Quine avance une hypothèse qui tient compte des rapports entre l'homme et son environnement : ce qui explique l'uniformité des valeurs morales les plus générales ou des fins ultimes, c'est l'uniformité des contraintes auxquelles est soumise l'action humaine. C'est une hypothèse assez banale (plutôt behavioriste). Elle ne mériterait peut-être pas qu'on s'y arrête si elle n'avait pas dans l'argumentation de Quine une fonction étonnante. Insister sur l'absence totale de continuité entre nos instincts et nos institutions morales, et insister parallèlement sur la continuité entre nos réactions naturelles et les constructions scientifiques permet à Quine d'expliquer pourquoi il est possible de *naturaliser l'épistémologie* et, par contraste, pourquoi il est impossible de *naturaliser l'éthique*. C'est sans doute une mauvaise nouvelle pour les réalistes moraux réductionnistes, mais les réalistes non réductionnistes ne se plaindront peut-être pas.

Il est exclu qu'on puisse débarrasser la théorie des valeurs morales de Quine de son behaviorisme (celui qui s'exprime dans

son insistance sur l'apprentissage). Mais il y a quelque chose à retenir du sentiment profond qu'il a de la discontinuité entre les valeurs morales et les « besoins » ou les préférences « naturelles », et de l'inanité des explications évolutionnistes ou fonctionnalistes de l'existence de ces valeurs.

Quoi qu'il en soit, Quine a estimé que son holisme ne pouvait pas être étendu à l'éthique : le holisme n'est pas un cohérentisme et ce que l'éthique peut offrir, au mieux, c'est un cohérentisme.

Certains philosophes estiment que Quine s'est platement contredit¹. D'après eux, on ne peut pas être à la fois holiste et partisan d'une stricte démarcation science-éthique. Ils estiment que la thèse de la démarcation est une réminiscence du positivisme logique que Quine avait pourtant complètement discrédité. Ils semblent penser que les positivistes logiques étaient dans l'erreur, mais que leurs thèses avaient le mérite d'être cohérentes et discutables alors que Quine est tout simplement incohérent.

Ces philosophes ont tort. L'argument holiste au sens strict n'implique pas l'effacement des frontières science-éthique, même s'il menace certaines versions de la dichotomie du fait et de la valeur. En définitive, le holisme, sur lequel certains réalistes moraux appuient leur théorie, ressemble plutôt à une sorte d'idéal moniste pour lequel Quine n'a guère de sympathie. Il est difficile de ne pas partager son manque d'enthousiasme.

1. Cf. Flanagan O.J., « Quinean Ethics », *Ethics*, 93, 1982, p. 56-74.

2. L'IMMUNITÉ DES RÉACTIONS MORALES

Comment parvenons-nous à concilier nos croyances (fortes) en la liberté, la responsabilité, la volonté et d'autres croyances tout aussi fortes mais apparemment incompatibles, en une certaine forme de déterminisme sociologique ou psychologique, déterminisme expliquant chacune de nos actions par notre « caractère » et les « circonstances » dans lesquelles nous avons l'habitude de nous retrouver, puis ce « caractère » et ces « circonstances » par notre éducation ou notre condition sociale ?

Strawson a proposé la solution suivante : nos jugements de réprobation ou d'approbation sont des réactions quasi-instinctives. De ce fait, elles ne peuvent pas être corrigées par nos croyances plus ou moins vagues dans le déterminisme sociologique ou psychologique, de la même façon que les illusions de la perception ne peuvent pas être corrigées par la connaissance de ce qui les provoque. C'est pourquoi tout comprendre, ce n'est jamais tout pardonner, et c'est heureux, car de cette « immunité » de nos réactions morales à l'égard de nos engagements déterministes, dépend la survie de la communauté¹.

Strawson nous dit que cette solution est inspirée, en grande partie, de considérations évoquées par Hume dans la section VIII de *l'Enquête sur l'entendement humain*, intitulée : « Liberté et nécessité ». Mais sur ce sujet au moins, les thèses du philosophe écossais sont loin d'être limpides. Hume devrait avoir certaines réserves à l'égard de la comparaison entre nos réactions morales authentiques et nos réactions perceptives, puisqu'il refuse d'assimiler les premières à des « sentiments naturels ».

Dans ces conditions, savoir pourquoi nos réactions morales authentiques demeurent à l'abri de nos engagements déterministes reste une question ouverte. Avant de revenir sur

1. Strawson P. F., « Freedom and Resentment » dans *Freedom and Resentment and Other Essays*, Londres, Methuen, 1974, p. 1-25.

cet argument, je vais essayer de préciser la question de Hume : celle de la compatibilité entre liberté et nécessité¹.

1

Le fait qu'une personne ait délibéré avant d'agir ou qu'elle ait planifié son action ne suffit certainement pas à prouver qu'elle ait été « libre » d'une manière importante quelconque. Car on peut toujours supposer que ses choix furent, pour ainsi dire, « contraints », « limités », « déterminés » par ses décisions précédentes. D'autre part, rien ne nous interdit de penser qu'il existe des manières habituelles de délibérer, c'est-à-dire une certaine continuité dans les choix, qui doit beaucoup à l'éducation reçue ou à toutes sortes d'autres contraintes psychologiques ou sociales.

Ces observations valent pour les comportements dits blâmables ou défectueux autant que pour les comportements dits corrects ou dignes d'éloge. Nous imputons le comportement dit correct aux bonnes habitudes, à la bonne éducation, etc. Si l'agent digne d'éloge est vraiment vertueux, il dira de lui-même qu'il n'est pour rien ou presque dans ses succès ou qu'il doit tout ou presque à ses parents, à ses maîtres, aux circonstances, etc. Car s'il déclare être pleinement responsable de ses succès, on pensera, à juste titre, qu'il est bien orgueilleux ou présomptueux. Mais si nous demandons de celui qui réussit qu'il reconnaisse sa part de chance (ou même qu'il doit tout à des formes de chance), pourquoi ne pas admettre que celui qui échoue doit tout à la malchance ? Même si nous acceptons une part de responsabilité dans la formation de notre caractère, celle-ci peut se réduire à presque rien lorsque nous ajoutons, pour expliquer un acte, les « circonstances ». Le chauffard qui s'est délibérément soûlé peut admettre la responsabilité de son ivresse tout en invoquant la malchance pour se disculper. Car le fait qu'un piéton traversait la chaussée

1. Voici les abréviations que j'utiliserai pour faire référence aux ouvrages de David Hume : T.N.H. : *Traité de la nature humaine* (1739-1740), trad. A. Leroy, Paris, Aubier, 1983 ; E.E.H. : *Enquête sur l'entendement humain* (1748), trad. M. Beyssade, Garnier-Flammarion, 1983 ; E.P.M. : *Enquête sur les principes de la morale* (1751), trad. A. Leroy, Paris, Aubier, 1947.

à l'instant précis où il grillait le feu rouge, par exemple, ne dépendait pas de lui, bien évidemment. Les chauffards ivres qui n'écrasent personne ne manquent pas, heureusement, la bonne fortune est de leur côté, tout simplement¹. On peut dissoudre graduellement la responsabilité des actes heureux ou défectueux en invoquant le caractère et les circonstances, qui sont des choses qui ne dépendent pas de nous et suffisent à expliquer nos actes.

Telle est, semble-t-il, la signification simple qu'on peut donner à ce fameux déterminisme sociologique ou psychologique, s'il en a une. Il s'en suivrait que tout ce que nous disons de la liberté, de la responsabilité, de la volonté serait illusoire et, à bien y réfléchir, totalement incompréhensible.

Or, si tout ce que nous disons de la liberté, de la responsabilité est illusoire, les idées de faute, de mérite, de blâme, de louange, de sanction en général deviennent artificielles, justifiées au mieux par leur utilité sociale, condamnées à disparaître, comme le suggérait Lévy-Bruhl à la suite de Comte, au fur et à mesure que notre connaissance objective du comportement humain ira en s'approfondissant : « Il est permis de prévoir, pour la nature morale, un processus analogue à celui qui s'est déroulé pour la nature physique : à mesure que la réalité morale deviendra davantage objet de science, elle sera moins objet de sentiment. L'œuvre d'Auguste Comte symbolise très exactement cette transition². »

Nous continuerons sans doute à punir, pour des raisons d'utilité, mais froidement, sans colère à l'égard des criminels, de même que nous asservissons la nature sans éprouver de sentiment à son égard (malheureusement, pourrait-on dire). Cependant, cette vision positiviste n'est pas nécessairement impliquée par la reconnaissance de la validité de la thèse déterministe.

Je vais me servir, à présent, d'un vocabulaire assez peu élégant, mais qui pourra nous aider, j'espère, à distinguer plusieurs autres possibilités. J'opposerai deux thèses : l'une dite incompatible et l'autre dite compatible. La thèse incompatible nous demande de choisir entre nos croyances dans le détermi-

1. Nagel T., *op. cit.*, 1983, p. 38-53.

2. Lévy-Bruhl L., *La morale et la science des mœurs*, Paris, P.U.F., 1971, p. 252-253.

nisme et nos croyances en la liberté, sous le prétexte qu'elles sont contradictoires. La seconde, dite compatibiliste, affirme qu'il n'est nullement besoin de choisir entre ces croyances, car elles ne sont pas contradictoires. Chacune de ces positions se subdivise de la façon suivante.

D'abord, l'incompatibiliste. Il peut dire : puisqu'il faut choisir, je choisis le déterminisme qui est plus convaincant que la liberté. Nos croyances en la liberté et la responsabilité, nos réactions affectives d'approbation et de désapprobation sont absurdes et devront progressivement céder la place à des vues plus raisonnables. Ce serait la position qu'on pourrait appeler « spinoziste » : ni rire ni pleurer, mais comprendre. C'est celle de Lévy-Bruhl, nous l'avons vu.

Mais l'incompatibiliste peut dire également : puisque je suis contraint de choisir, eh bien, je choisis la liberté, qui est moins indéfendable que le déterminisme. L'incompatibiliste « libertaire », comme on peut l'appeler, dira par exemple : l'idée que nous aurions pu faire autre chose que ce que nous avons fait n'est pas inintelligible. Il n'est pas vrai que notre caractère, aussi pervers soit-il, et les circonstances, aussi défavorables soient-elles, entraînent nécessairement tel ou tel acte. Les cas dans lesquels le même caractère et les mêmes circonstances n'entraînent pas le même acte défectueux ne manquent pas ! Et nous n'avons aucune bonne raison de penser que cela est dû tout simplement au fait que nous avons mal identifié les caractères et les circonstances. Je n'évoquerai pas, ici, les autres manières, plus subtiles (sartriennes, par exemple) de justifier le choix de la liberté (s'il faut choisir, bien évidemment, ou si on est libre de choisir, ce que le déterministe nierait, bien entendu).

Voyons maintenant ce que dit le compatibiliste. Pour lui, rien ne nous oblige à choisir entre la liberté et le déterminisme car ces thèses ne sont pas contradictoires. Mais à nouveau, cette opinion se subdivise, car on peut être compatibiliste par inclusion ou compatibiliste par exclusion.

Le compatibiliste par inclusion affirmera : tout ce que nous disons de la liberté est analytiquement lié à l'idée de nécessité. Si nous ne pensions pas que toutes nos actions sont rigoureusement déterminées par des motifs, nous ne pourrions pas du tout affirmer qu'elles sont libres. Nous ne disons jamais d'un fou imprévi-

sible qu'il est libre. C'est précisément la régularité de notre comportement qui est l'indice de sa liberté.

De son côté, le compatibiliste par exclusion soutiendra : ce que nous disons de la nécessité ou du déterminisme porte sur un domaine entièrement séparé de ce que nous disons de la liberté, de la responsabilité ou de la volonté. Comme ces deux thèses ne se rapportent pas du tout à la même chose, elles ne peuvent pas se contredire. Notre compréhension des actes humains est une chose ; nos réactions d'indignation ou d'approbation en sont une autre. D'ailleurs, tout comprendre, ce n'est pas tout pardonner, on le sait bien.

Dans la section VIII de *l'Enquête sur l'entendement humain*, David Hume passe du premier compatibilisme, par inclusion, au second, par exclusion. Entre ces deux compatibilismes, il y a certainement une tension. Mais, pour Hume, ils se rejoignent au fond, car ce sont des échos de ses attaques dévastatrices contre le réalisme, l'objectivisme, le rationalisme en général, dans le domaine moral. Il ne faut pas s'en étonner, car l'empirisme de Hume, qui devrait, en principe, lui interdire d'ériger un système ne l'empêche tout de même pas de donner une certaine cohérence à ses thèmes. La théorie de la connaissance de Hume nous donne certaines clés indispensables pour comprendre sa doctrine morale et clarifier la question du compatibilisme.

2

L'empiriste semble se demander : comment puis-je passer de la phrase « Je vois une rose rouge » à la phrase « Il y a une rose rouge » ? Pour l'empiriste, entre ces deux énoncés, il y a un gouffre infranchissable : « Je vois une rose rouge » est un compte rendu de sensations et, à ce titre, il offre certaines garanties d'évidence ou d'incorrigibilité. Car de même que je ne peux pas douter, à première vue, de mon plaisir ou de ma douleur, je ne peux pas douter de mes sensations visuelles. Le doute ne peut porter que sur la relation entre ces sensations et ce qu'elles sont censées représenter ou entre ces sensations et les raisons que j'ai de les avoir.

La proposition : « Il y a une rose rouge » n'est pas un compte rendu de sensations. Elle suppose la référence à un objet permanent et indépendant, ce qu'aucune sensation n'est en mesure de

garantir. Nous ne pouvons pas justifier notre croyance en l'existence d'objets ayant une existence permanente et indépendante de nos sensations à partir de comptes rendus de sensations. Cette justification peut-elle être logique ? Non. La proposition « Il y a une rose rouge » ne peut pas être déduite de « Je vois une rose rouge ».

Pour expliquer ce passage qui n'est justifié ni par la logique, ni par nos sensations, Hume ne conçoit pas d'autre possibilité que l'intervention d'un jeu purement imaginaire. La rose rouge en tant qu'objet permanent et indépendant est une sorte de fiction, ou plus exactement une croyance que nous acceptons habituellement ou spontanément, mais elle n'a aucune garantie logique ou sensible.

Les conclusions de Hume sont donc étroitement liées à sa conception de la structure de l'entendement humain qui serait composé de trois facultés : celle de raisonner logiquement sans le secours de l'expérience, celle de percevoir, et une faculté d'imagination. Si quelque chose ne dépend ni de la raison ni des sensations, alors cette chose ne peut être liée qu'à l'imagination.

Dans son analyse du jugement moral, Hume applique ces principes. Si nous pensons qu'un acte est vicieux ou vertueux, ce ne peut être en vertu d'une relation logique, car celle-ci ne fait pas de distinction entre, par exemple, le parricide et le remplacement d'un chêne par ses propres graines (*T.N.H.*, p. 582-583). En disant qu'un acte est vicieux ou vertueux, nous utilisons donc nos deux autres facultés. Nous éprouvons quelque chose, un sentiment d'approbation ou de désapprobation, par le jeu de nos facultés de sentir et nous projetons ces sensations, pour ainsi dire, dans l'acte, par le jeu de notre imagination. Il en résulte que ce que nous éprouvons et projetons, nous finissons pas l'attribuer à l'acte même.

Nous croyons que les propriétés « être bon », « mauvais », « courageux », « cruel », « généreux », sont les propriétés objectives d'un objet indépendant (un acte, une personne), alors que ce sont des sensations (comme les saveurs, les sons, les couleurs, les odeurs) et des projections de l'imagination (*T.N.H.*, p. 584-585). Je n'insisterai pas, ici, sur les limites de cette méthode qu'on continue d'appliquer lorsqu'on essaie par exemple de rapporter toutes choses au symbolique, au réel et à l'imaginaire, par

exemple. Si quelque chose n'est pas lié à deux de ces idées, alors elle l'est nécessairement à la troisième, et voilà tout le raisonnement.

Dans sa théorie de la connaissance, Hume prétend partir du certain, c'est-à-dire ou bien de la logique ou bien de nos sensations, et il se demande jusqu'où nous pouvons arriver avec ces bagages. Et sa réponse est claire : certainement pas jusqu'aux objets extérieurs, ni même au moi, en tant qu'objet identique à lui-même dans le temps. Kant disait de Hume qu'il était un géographe de la raison humaine¹. Mais, sans diminuer les mérites de cette formule, on pourrait dire aussi qu'il en est l'historien, ou plutôt qu'il élabore une histoire naturelle de la raison humaine.

Cette histoire commence avec nos impressions, notion extrêmement vague qui comprend tout ce qui est événement mental, sensations liées à l'audition, la vue, le tact, mais aussi sentiments de plaisir ou de peine. Ces impressions ont des objets, mais ce ne sont pas ces objets qui les causent. La direction va des impressions au monde et non du monde aux impressions.

Ce que Hume semble nous demander d'accepter, c'est que tout commence par des impressions, un flux ininterrompu d'événements mentaux, les uns vifs et les autres plus faibles. À ce moment, il n'y a pas encore de moi, ni même de sujet percevant, ni d'objet extérieur. D'emblée, Hume attaque notre ontologie ordinaire peuplée de personnes, d'objets, et de représentations d'objets dans l'esprit des personnes. Ce flux ininterrompu d'événements mentaux s'organise cependant, mais sans le détour d'un « moi ». Nos idées s'associent d'elles-mêmes, comme si elles vivaient dans un univers platonicien, autonome ; elles s'accordent, se juxtaposent, s'excluent. Selon Hume, il n'est pas absurde, ou logiquement contradictoire, d'affirmer qu'elles existent sans sujet percevant : ce sont des événements antérieurs aux substances. Cette vie des idées a peut-être un substrat physique pour lieu : mais ce lieu n'en est que le réceptacle, le « théâtre ». Ce n'est pas un « moi » souverain, maître du processus. Jusqu'à présent, il n'y a qu'un « amas », une « collection »

1. Kant E., *Critique de la raison pure* (1787), trad. A. Tremesaygues et B. Pacaud, Paris, P.U.F., 1975, p. 519.

d'impressions (*T.N.H.*, p. 296). Et Hume ajoute : « Il n'y a pas d'absurdité à séparer de l'esprit une perception particulière ; c'est-à-dire à rompre tous ses rapports avec cette masse de perceptions conjointes qui constituent un être pensant » (*T.N.H.*, p. 296). Si cette thèse était tellement absurde d'ailleurs, elle n'aurait peut-être pas survécu jusqu'à nos jours où elle prend la forme du « fonctionnalisme » en psychologie (doctrine insistant, on l'a vu, sur l'indépendance des événements mentaux en tant que types à l'égard des substrats physiques).

Hume s'inspire de Newton pour suggérer que les idées s'associent entre elles selon quelques principes simples. Ces principes peuvent être ramenés à trois : a) ressemblance ; b) contigüité dans le temps et l'espace ; c) relation de cause à effet.

Hume illustre ces relations de la façon suivante : « Un tableau conduit naturellement nos pensées à l'original [ressemblance], la mention d'un appartement dans une maison introduit naturellement une enquête sur les autres appartements [contigüité]. Et si nous pensons à une blessure, nous pouvons à peine nous empêcher de réfléchir à la douleur qui suit [cause et effet] ». (*E.E.H.*, p. 72.)

On peut observer que ces relations de ressemblance, contigüité cause et effet ne sont nullement logiques (en logique, on parlerait d'identité, conjonction, implication ; mais la comparaison est inappropriée en réalité).

De ces trois relations, c'est à celle de cause et d'effet que Hume accorde le plus d'importance. C'est sur elle qu'il porte sa fameuse attaque. Selon Hume, c'est cette relation qui nous permet en quelque sorte d'engendrer un monde ayant des caractères stables, permanents, ordonnés.

Parallèlement à cette genèse du monde physique à partir de nos impressions de sensation, on peut parler de la genèse d'un monde moral à partir d'une sous-catégorie d'impressions : les passions (colère et joie, honte et fierté, haine et amour, etc.).

Hume ne souhaite pas distinguer trop nettement impressions de sensations et sentiments de plaisir ou de peine. Il appelle passions les impressions liées à ces sentiments. Cependant, l'histoire naturelle des impressions liées aux sens externes, ouïe, vue, etc. n'est pas identique à celle des passions. En effet, l'impression de sensation est primitive. Rien ne la précède dans l'esprit, elle s'édulcore en idée, grâce à la mémoire. Mais les sentiments de

plaisir ou de peine peuvent être éveillés par quelque chose qui les précède dans l'esprit. Rien, dans l'esprit, ne précède une perception de rose rouge, mais un chagrin peut être précédé par l'idée d'un deuil, dans l'esprit. C'est pourquoi Hume préfère appeler les impressions liées à des idées, impressions de réflexion.

On voit donc comment ces histoires naturelles se distinguent. Hume suggère qu'on passe naturellement d'une impression à l'idée correspondante par affaiblissement, alors que, dans les passions, on passe d'une idée à une impression forte, ce qui n'est possible, dans le cas des impressions de sensation, que lorsque nous sommes hallucinés. Cependant, dans les impressions de réflexion il convient plutôt de parler d'une combinaison d'idées et de sentiments : l'idée d'être ruiné et le déplaisir que cette idée provoque peuvent être considérés comme des données primitives.

Les lois d'association (ressemblance, contiguïté, causalité) opèrent pour engendrer l'idée du moi et d'autrui. De quelle façon ? Certaines idées associées à certains objets causent la honte ou la fierté, et celles-ci se donnent le moi comme objet. De la même façon, certaines idées associées à certains objets causent la haine ou l'amour, et celles-ci se donnent autrui comme objet. Pour Hume, il n'y a pas de moi qui puisse « avoir honte » ou « être fier », il y a d'abord des sentiments de honte ou de fierté qui me font croire au « moi ». De la même façon, il n'y a pas d'autrui qui provoque ma haine ou mon amour. C'est ma haine ou mon amour qui engendre l'idée d'autrui. Dans ces engendrements, l'imagination remplit un rôle éminent : c'est elle qui autorise ces projections du moi et d'autrui comme elle autorisait la constitution d'objets continus et permanents.

Je sais qu'il n'est pas très facile d'accepter cette inversion de notre ontologie ordinaire dans laquelle nous avons d'abord des personnes et ensuite des sentiments et des propriétés. D'ailleurs, la thèse de Hume est parfois confuse sur ce point, mais il me semble que c'est bien de cette façon qu'il convient de la comprendre, en harmonie avec ce qu'il nous dit de la constitution des objets par l'imagination. En tout cas, elle annonce bien des doctrines modernes sur la genèse du moi (le moi ne serait pas « donné » mais « construit » psychologiquement ou socialement, etc.).

Ce que nous pouvons retenir de cette histoire, c'est l'élimination de l'idée de représentation, des privilèges du moi, du monde et l'importance de l'imagination dans la formation d'objets indépendants et permanents. Pour mieux comprendre ce dernier point, il faut revenir un peu sur la théorie de la causalité de David Hume.

La section IV de l'Enquête sur l'entendement humain s'ouvre par cette affirmation très forte : « Tous les objets de la raison humaine ou de nos recherches peuvent naturellement se diviser en deux genres, à savoir les relations d'idées et les faits » (*E.E.H.*, p. 85). Il est difficile de ne pas voir dans cette proposition une anticipation de la distinction que Kant rendra fameuse, entre propositions analytiques et synthétiques. Mais le point est très controversé. Pour Hume, le critère d'une relation de fait, c'est que nous n'avons aucune difficulté à concevoir sa négation. « Le soleil ne se lèvera pas demain », n'est pas moins intelligible que « Le soleil se lèvera demain ». Alors que la négation de « Trois fois cinq est égal à la moitié de trente » est inconcevable, la négation de « Le soleil se lèvera demain » n'est pas du tout inconcevable. Les relations de fait ne sont pas des relations logiques et c'est d'ailleurs pour cette raison qu'elles ne peuvent pas être conçues *a priori*, dans la tranquillité d'une réflexion philosophique.

D'après Hume, la relation de cause à effet ne peut pas être conçue *a priori*. Peut-on dire alors qu'elle est un fait d'observation ? Hume rejette cette hypothèse. Soit le cas d'une boule de billard qui, par son impact, semble entraîner le déplacement d'une autre boule de billard. Qu'observons-nous en réalité ? Le mouvement de la première, un bref contact entre les deux boules et le mouvement de la seconde. Mais nous n'avons pas « vu » la causalité. Nous ne pourrions prétendre avoir saisi cette relation que si nous avons des raisons de penser que le déplacement de la deuxième boule n'est pas un miracle, l'effet de la volonté divine ou d'un mécanisme caché sous la table. En fait, il faudrait donner une traduction contrefactuelle de la proposition « Le mouvement de la première boule a causé le déplacement de la seconde » et dire : « Si la première boule n'avait pas frappé la seconde, le seconde ne serait pas déplacée ».

Ainsi, dans la notion complexe de causalité, nous pouvons isoler trois idées : priorité (il y a un antécédent et un conséquent) ; contiguïté (il y a un contact direct entre deux objets) ; connexion nécessaire. Il n'est pas facile de donner une signification simple à cette dernière idée. En un sens, c'est une illusion. Voyons pourquoi.

Pour Hume, l'ordre des séquences temporelles et la contiguïté ne suffisent pas à décrire une relation causale. Il faut y ajouter celle d'une connexion dont la nécessité s'exprime sur le mode contrefactuel. S'agit-il d'une relation logique ? Nous avons vu que tel n'était pas le cas. Hume dit qu'il n'y a rien d'« illogique » dans l'affirmation « Tous les arbres fleuriront en décembre et dépériront en mai », ou dans « Un corps tombant des nuages et qui à tous autres égards ressemble à de la neige a pourtant la saveur du sel et le contact du feu » (*E.E.H.*, p. 95). D'un point de vue logique, n'importe quoi peut sortir de n'importe quoi. Si la relation de causalité n'est pas observable parce qu'elle va au-delà du donné, et si elle n'est pas une relation logique car sa négation n'implique pas contradiction, de quel ordre est-elle ? En d'autres termes : si cette relation n'est ni impression, ni acte de l'entendement, d'où peut-elle venir ? Pour Hume, nous n'avons pas d'autre possibilité que l'attribuer à notre imagination. En réalité, il n'y a pas de connexion nécessaire, mais il se forme dans notre esprit imaginatif, par des lois d'association, lorsque nous percevons tel ou tel événement, l'attente de l'apparition d'un autre événement. Ainsi, en voyant une flamme, nous nous attendons à ce qu'elle brûle la feuille de papier que nous posons sur elle. D'où vient cette attente ? Pour Hume, elle a pour seule origine le souvenir de la succession de ces événements. Et, désormais, lorsque nous sommes en présence du premier, le contact du papier et de la flamme, nous attendons le second, la destruction du papier par les flammes. Mais pourquoi l'avenir devrait-il ressembler au passé ? Hume répond : parce que l'imagination nous entraîne à croire très vivement à l'existence de lois de la nature. Mais, en réalité, nous sommes seulement accoutumés à ce que le papier brûle lorsqu'il est mis au contact d'une flamme. C'est l'habitude qui commande. Mais rien ne garantit logiquement, démonstrativement, c'est-à-dire certainement, qu'il en ira toujours ainsi. Nous prédisons parfois un événement sans avoir eu l'expérience d'une association. Mais ce

n'est qu'une illusion. Car en analysant ce que nous faisons alors, nous nous apercevons vite que nous réduisons cette prédiction à des successions déjà connues de nous.

Ainsi les prétendues lois de la nature ne sont qu'une projection de nos habitudes. Il n'y a pas de lois de la nature aussi fermes que les lois logiques. Il n'y a que des reconstructions apparemment démonstratives à partir de nos expériences passées. Et c'est ainsi que nous succombons à l'illusion de la connexion nécessaire. Ces nécessités sont en nous, dans nos attentes des répétitions de ce que nous avons toujours observé. L'idée d'une loi de la nature est une sorte de postulat indémontrable, même s'il est fructueux et utile.

Il me semble que nous accepterons plus aisément le renversement que Hume nous propose lorsqu'il nous demande de nous interroger sur la signification des prédicats de sensation : être bleu, rouge, chaud, lisse, rugueux. Il ne nous est pas trop difficile d'admettre que ce que nous croyons être une propriété intrinsèque de l'objet n'est, en réalité, qu'une perception de l'esprit ou, au mieux, une relation entre certaines propriétés de l'objet et notre constitution naturelle. Nous acceptons plus facilement encore l'idée que les douleurs physiques sont en nous plutôt que dans l'objet. Lorsque je suis piqué par une épingle, je ne dis pas que « l'épingle est dolorifique », alors que j'ai tendance incontestablement à dire de mes sensations de rouge que ce sont des perceptions des propriétés de l'objet rouge. Enfin, lorsque nous parlons de propriétés esthétiques, nous admettons sans difficultés que rien, peut-être, dans le visage d'une personne aimée ne détermine certainement notre sentiment de sa beauté. Nous projetons cette beauté. Elle est en nous, non dans l'objet.

À présent, nous avons assez d'éléments, il faut l'espérer, pour comprendre de quelle façon Hume entend défendre parallèlement les deux formes de compatibilité entre liberté et nécessité, dont on pourrait croire, à première vue, qu'elles sont incompatibles. Mais nous avons des indications, aussi, sur ce qui pourrait justifier la thèse d'après laquelle il aurait échoué¹.

1. Boatright J.R., « Hume's Account of Moral Sentiment », *Revue internationale de philosophie*, 115-116, 1976, p. 79-90.

Si nous admettons, avec Hume, que la prétendue nécessité se résume à l'observation de régularités ainsi qu'à l'attente de leur répétition, nous ne songerons plus à dissocier le règne de la nature et le règne de la société. C'est parce que nous nous faisons une idée erronée, peut-être trop forte, de la nécessité dans la nature que nous répugnons à concevoir la nécessité dans la société. Qu'est-ce qui se joint dans le domaine pratique et quelles sont nos attentes ? Quels sont les objets semblables qui s'associent régulièrement ? Ce sont des caractères, des motifs, des circonstances et des conduites. Nous comptons que le même caractère, le même motif, la même circonstance entraîneront la même conduite.

« Si un homme que je connais pour honnête et riche et avec qui je vis en amitié intime était pour venir chez moi où je suis entouré de mes serviteurs, je reste persuadé qu'il ne me frappera pas de son poignard avant de partir pour me dérober mon encrion d'argent ; je ne redoute pas plus cet événement que la chute de la maison elle-même, qui est nouvelle et bien bâtie sur de bonnes fondations. » (*E.E.H.*, p. 160.)

C'est un ami bon et honnête (un caractère) qui rend une visite amicale (un motif) dans une maison où l'hôte vit entouré de serviteurs (une circonstance) ; tout ceci permet d'attendre raisonnablement un rendez-vous paisible. Toute notre vie pratique repose sur cette confiance, ces attentes, ces croyances dans la répétition.

« Le plus pauvre artisan, qui travaille seul, compte au moins sur la protection du magistrat qui lui garantit les fruits de son travail. Il compte aussi que, lorsqu'il portera ses marchandises au marché, et qu'il les offrira à un prix raisonnable, il trouvera des acquéreurs et qu'il aura le pouvoir, grâce à l'argent qu'il aura obtenu, d'engager les autres hommes à lui fournir les commodités nécessaires à sa subsistance. À mesure que les hommes étendent leurs transactions et qu'ils compliquent davantage leurs relations avec les autres hommes, ils comprennent toujours, dans leurs programmes de vie, une plus grande diversité d'actions volontaires qu'ils comptent, pour de justes motifs, voir coopérer avec leur action propre. Pour toutes ces conclusions, ils tirent leurs règles de l'expérience passée, de même manière que

pour leurs raisonnements sur les objets extérieurs ; et ils croient fermement que les hommes, aussi bien que les éléments, doivent rester, dans leurs actions, les mêmes qu'ils les ont toujours trouvés. » (E.E.H., p. 158.)

Nous comptons sur, nous attendons, nous croyons fermement dans la répétition ; c'est l'évidence de notre vie pratique. Mais cela signifie-t-il que ces attentes, cette confiance, ces croyances soient, en quelque sorte, soutenues par le « raisonnement expérimental » ? (E.E.H., p. 158.) Comment pouvons-nous établir le fait que, dans notre vie pratique, nous avons observé des conjonctions constantes en même temps que nous formions ces habitudes qui nous inclinent à attendre la répétition de ces conjonctions ?

L'argument de Hume repose sur l'évidence de nos attentes. Quant à l'observation de conjonctions constantes, elle est prouvée par le fait que nous attendons. C'est une démonstration circulaire. Nous avons des attentes, c'est un fait. Si nous avons des attentes, c'est qu'il existe des régularités observables ; comment savons-nous qu'il existe des régularités observables ? Parce que nous avons des attentes. Hume considère sans doute que ces attentes disparaîtraient si elles étaient démenties par les faits. C'est le sens de son allusion au « raisonnement expérimental ». Mais comment être assurés que nous ne pourrions pas, dans le domaine pratique, toujours attendre ce qui est toujours démenti ? Les illusions amoureuses ne suffisent-elles pas à démentir la force du raisonnement expérimental ?

Nous ne manquons pas de prendre nos désirs pour des réalités, et ce mouvement irrationnel assez commun devrait affaiblir l'équivalence entre l'attente et la conjonction constante.

Toutefois, Hume ne s'adresse pas cette objection. La seule incertitude qu'il entrevoit ne modifie pas l'identité qu'il postule entre règne de la nature et règne de la société. Si l'attente est démentie, ce n'est pas que telle est la vocation essentielle de l'attente ; c'est que la réalisation de ce qui est attendu est probable mais non certaine. Supposons en effet que le bon ami riche et honnête soit pris d'une folie « subite et inconnue ». Ce n'est pas inconcevable. Mais tout ce qu'on pourra en dire, c'est que l'occurrence de la conduite attendue doit « être proportionnée à notre expérience de la conduite habituelle des hommes dans de telles situations particulières » (E.E.H., p. 160). La probabilité peut

être très forte sans jamais atteindre la certitude. Et Hume demande : nos inférences dans le domaine de la nature ne sont-elles pas limitées de la même manière ? Il n'est pas utile, ici, d'approfondir la discussion de cette adjonction. Hume parvient à sauver, provisoirement, son analogie entre règne de la nature et règne de la société. Puisque nous comptons sur une certaine constance dans les actions humaines, et puisque ces attentes ne peuvent être fondées que sur l'observation de régularités, notre vie pratique est gouvernée par des conjonctions constantes et réglée par l'habitude. C'est dire qu'elle est soumise à la nécessité. S'ensuit-il qu'il n'y ait plus de place pour le mérite et la culpabilité, la louange et le blâme, la justification du jugement moral ? C'est tout le contraire qu'il faut penser. Sans idée de nécessité, nous ne pourrions concevoir ni le blâme ni la louange, ni le mérite et ni la culpabilité.

Tout d'abord, ce qu'on blâme, ce n'est jamais l'action elle-même, qui est incorporelle, évanescence, périssable (*E.E.H.*, p. 167). Pour saisir l'argument de Hume, il faudrait peut-être se demander ce que voudrait dire punir le crime sans punir le criminel. Ce serait une absurdité. L'action doit être assignée causalement à un corporel pour que le blâme ait une signification. De quelle nature doit être ce corporel ? Ce ne peut être un objet inanimé bien entendu : les crimes d'un chêne ou d'un orme nous laissent indifférents. Ce ne sont d'ailleurs pas du tout des crimes, du fait même qu'ils nous laissent indifférents.

Ce qui peut exciter nos sentiments et par conséquent susciter l'éloge ou le blâme, c'est seulement la personne, la créature douée de pensée et de conscience : « Il est évident d'abord que le seul objet propre de la haine ou de la vengeance est une personne ou une créature douée de pensée et de conscience, et quand les actions injurieuses excitent cette passion, c'est seulement par leur relation ou leur connexion avec elles. » (*E.E.H.*, p. 167.)

Mais qui dit « personne » dit aussi « constance ». Du moins nous devons croire en cette constance, sans quoi il n'y aurait aucun sens à parler de mérite ou de culpabilité. Si la personne qu'on entend punir n'était pas la même que la personne ayant commis la faute, comment pourrions-nous la blâmer ? On ne blâme pas les hommes pour des actions qu'ils accomplissent à leur insu et par accident ? On blâme moins ceux qui ont eu un écart de

conduite qui semble passer ? C'est qu'il ne s'agit pas tout à fait de la même personne, du caractère dans sa constance. C'est lui et lui seulement qui suscite nos réactions morales (*E.E.H.*, p. 167). Notre croyance dans la constance du caractère se manifeste aussi dans notre croyance à l'uniformité des effets de certains motifs (punitions, récompenses). Il n'y aurait aucun sens à élaborer des systèmes de punitions, de lois, de récompenses, si on n'admettait pas comme principe fondamental que ces motifs, être récompensé, être puni, ont une influence régulière et uniforme sur l'agent, « produisent les bonnes actions et préviennent les mauvaises » (*E.E.H.*, p. 166). Cette inférence, il est difficile de ne pas la concevoir comme une cause et de ne pas la considérer comme un exemple de la nécessité.

Ainsi, nous serions incapables de donner un sens au jugement moral si nous ne pouvions dépasser le donné, croire, inférer, attendre, compter sur la répétition de conjonctions constantes. Bref, si le jugement moral n'était pas entièrement soutenu par notre idée de la nécessité, il ne signifierait rien pour nous.

Hume ne s'intéresse pas à nos façons d'assigner une responsabilité, mais aux conditions de possibilité d'un jugement moral en général. Son argument n'est pas aussi psychologique ou génétique qu'on le dit habituellement. C'est d'ailleurs la forme de son argument qui invalide l'objection assez courante faite à son traitement de l'idée de cause. On a reproché à Hume d'avoir fait reposer l'idée de cause sur celle de conjonction constante. N'existe-t-il pas des cas d'assignation directe de causalité ? Nous pouvons fort bien concevoir directement la causalité, sans inférence, par une sorte d'obligation conceptuelle liée à notre usage des verbes « tuer », « persuader », « construire », etc. Par ailleurs, l'imputation d'une responsabilité ne ressemble-t-elle pas à la perception d'un tracé, observable comme les marques du parcours d'un objet d'un point vers un autre ? On pourrait donc concevoir directement la causalité ou même l'observer, c'est-à-dire nous servir de l'entendement et des sensations sans avoir à utiliser l'argument coûteux de l'habitude et de l'imagination. Mais l'exemple de l'assignation de responsabilité n'est pas convaincant. Supposons que nous souhaitions répondre à la question : « Qui a volé le crayon qui était posé sur la table ? » Parmi les réponses acceptables, nous n'incluons probablement

pas : c'est le taille-crayon (un objet inanimé ne « vole » rien), ou je l'ai fait disparaître de mes pensées (un événement intérieur ne « vole » rien). En réalité, nous allons orienter nos recherches vers une « personne », vers un caractère, c'est-à-dire que nous allons dépasser le donné. Nous n'imputons aucune « responsabilité » à un objet inanimé ou à un événement intérieur : l'opération de notre esprit, lorsque nous imputons une responsabilité, est une inférence.

D'ailleurs, toutes nos idées de responsabilité sont soutenues, d'une façon ou d'une autre, par des inférences car toutes dérivent de l'idée de liberté ; et celle-ci est inconcevable en dehors de la nécessité. Non pas bien sûr la liberté d'indifférence (mon acte n'a pas de motif contraignant, il est sans cause nécessaire, j'aurais toujours pu faire autre chose. C'est une pure spéculation : comment prouver que j'aurais pu faire autre chose ?) ; mais la liberté qui nous est donnée en acte (je fais ce pourquoi j'ai le motif le plus fort, ce qui correspond à ma nature, ce qui ne contrarie pas mon mouvement spontané). Cette liberté « conditionnelle » appartient à tout homme qui n'est pas prisonnier dans les chaînes (*E.E.H.*, p.164). Elle serait inconcevable si on ne pouvait pas établir une démarcation entre ce qui est de notre fait et ce qui ne l'est pas, c'est-à-dire entre ce qui dépend du caractère intérieur et ce qui n'en dépend pas. Mais pour avoir l'idée d'un caractère intérieur, il faut concevoir la constance, et, encore une fois, la nécessité.

De quelque côté qu'on prenne l'idée de liberté, elle est imprégnée de nécessité. La nécessité garantit la possibilité d'une constance de la personne et, par conséquent, de ce qui peut dépendre d'elle ; elle garantit aussi la constance des conduites, l'inclination à suivre des motifs. C'est à ces formes constantes qu'on reconnaît la liberté, qui n'est certainement pas l'expression d'un tourbillon de modifications de la personne, de conduites aléatoires, d'actions folles réglées par le hasard.

Ainsi, liberté et nécessité sont bien compatibles puisque la doctrine de la liberté dépend de la nécessité. Ce qu'on appelle liberté n'est, à proprement parler, qu'une variété de la nécessité. C'est cette façon d'inclure la liberté dans la nécessité qu'on a coutume d'appeler le compatibilisme par inclusion de Hume.

Mais ce compatibilisme par inclusion ne permet pas véritablement de dégager la spécificité du jugement moral. Tout ce que Hume a montré, c'est que nos idées de liberté ou de responsabilité, d'éloge ou de blâme n'ont de sens que si nous admettons le règne de la nécessité dans les affaires humaines. Mais, dans ces conditions, une boule de billard pourrait être dite « libre ». Le seul argument que Hume avance pour exclure la possibilité de condamner une boule de billard qui en frappe une autre, c'est que nos réactions d'éloge et de blâme ne peuvent avoir pour objets que des personnes. Mais, qu'est-ce qui justifie cette restriction ? En réalité, le compatibilisme par inclusion ne nous permet pas de répondre à cette question. Hume semble penser qu'il est possible de remédier à cette insuffisance manifeste en associant au compatibilisme par inclusion, l'autre forme de compatibilisme, par exclusion. Pour justifier son point de vue, il n'hésite pas à se servir de ces remarques sarcastiques dont il est toujours très généreux. Il nous demande, en gros, si un homme torturé par des douleurs épouvantables serait apaisé si on essayait de le consoler en lui disant que ses souffrances, aussi profondes soient-elles, contribuent à l'harmonie universelle ou au plus grand bonheur possible de tous les êtres créés, ou que ses souffrances sont de médiocre importance au regard de sa place ridicule dans l'univers ou de toutes autres souffrances plus fortes, passées et présentes, endurées par d'autres hommes.

Dans un texte de l'*Enquête*, il prend le parti du bon sens en suggérant que ces consolations risqueraient fort de provoquer l'effet contraire de celui qui était attendu. Les remèdes des stoïciens, par exemple, seraient certainement mal accueillis. « Vous irriteriez sûrement plutôt que de l'apaiser un homme torturé par les douleurs de la goutte en lui exaltant la justesse des lois générales qui ont produit les humeurs malignes de son corps et les ont conduites par les canaux voulus, jusqu'aux tendons et aux nerfs où elles éveillent maintenant des tourments aussi aigus. » (*E.E.H.*, p. 170.) Ce qui vaut pour la douleur physique vaut aussi pour la douleur morale : « Un homme, à qui on dérobe une somme considérable trouve-t-il que la contrariété qu'il éprouve de cette perte soit en rien diminuée par ces réflexions sublimes ? »

(*E.E.H.*, p.171.) Dans l'*Enquête sur les principes de la morale*, Hume multiplie ce genre d'exemples, en affermissant de cette façon la comparaison entre jugement moral et perception. « Un homme peut aussi bien prétendre se guérir de l'amour en regardant sa maîtresse au moyen d'un microscope et de sa perspective artificielle, et en découvrant ainsi la grossièreté de sa peau et la monstrueuse disproportion de ses traits qu'espérer exciter ou modérer une passion à l'aide des arguments artificiels d'un Sénèque ou d'un Épicète. » (*E.P.M.*, p. 247.)

En fait, toutes ces considérations sont parallèles à celles que Hume fait au sujet du jugement esthétique, qui est entièrement indépendant de ce que la physique nous apprend. Dans les deux cas, nos réactions de plaisir ou de peine sont immunisées à l'égard de nos connaissances ou de nos croyances générales : « Le teint vif et agréable d'un visage ne m'inspire-t-il pas du contentement et du plaisir même si la philosophie m'apprend que toute la différence de ce teint naît de très petites différences d'épaisseur dans les plus petites parties de la peau ; et que telle est la cause qui permet à une surface de réfléchir l'une des couleurs primitives de la lumière et d'absorber toutes les autres ? » (*E.P.M.*, p. 158.)

Ainsi, Hume nous invite à admettre que nos réactions de plaisir ou de peine, qu'elles soient physiques, esthétiques, morales, sont pour ainsi dire, à l'abri des spéculations théoriques générales. Cette suggestion pourrait valoir pour la relation entre liberté et nécessité. Si la nécessité est purement théorique, et si la liberté peut être réduite, par sa traduction en termes d'éloge ou de blâme, à une réaction physique ou esthétique, alors le fameux problème de la liberté et de la nécessité serait aussitôt dissous.

Une analogie et une thèse suffiraient à justifier nos jugements moraux d'éloge et de blâme ; analogie de la douleur physique et de la douleur morale, thèse postulant l'étanchéité de nos raisonnements théoriques d'une part et de nos réactions physiques ou esthétiques de l'autre. C'est bien la thèse que Hume semble défendre lorsqu'il raille l'idée de la Providence, lorsqu'il ironise sur les thèmes sublimes des anciens stoïciens, lorsqu'il se moque de tous les docteurs Pangloss, avec Voltaire.

Hume résume cette thèse en une sèche formule : « Les sentiments naturels de l'esprit humain ne peuvent être commandés ni

modifiés par aucune théorie philosophique, ni par aucune spéculation que ce soit. » (*E.E.H.*, pp. 171-172.) Mais cette formule, Hume s'est permis de l'avancer après avoir réussi à repousser des objections considérables et en s'appuyant sur toute sa doctrine de la connaissance.

On peut se demander tout d'abord si David Hume ne se contredit pas en rejetant ces consolations stoïciennes pour leur évidente inefficacité, alors qu'il déclare par ailleurs que la raison, bien qu'esclave des passions, oriente tout de même ces passions. Nos désirs ne sont jamais purs ; ils s'accompagnent toujours d'une certaine idée. « Un homme transporté de joie par la possession d'un diamant ne se borne pas à regarder la pierre étincelante qui est devant lui. Il en considère aussi la rareté et telle est la principale cause de son plaisir et de son excitation. Ici donc un philosophe peut intervenir ; il peut nous suggérer des vues particulières, des considérations et des circonstances qui, autrement, nous auraient échappé et, par ce moyen, il peut soit modérer, soit exciter une passion particulière. » (*E.P.M.*, pp. 246-247.)

Par ailleurs, une passion telle que l'espoir ou la crainte, le chagrin ou la joie, le désespoir et la confiance, se fonde sur une supposition d'existence de l'objet qui peut satisfaire ces passions. Et quand il est montré ou démontré que cet objet n'existe pas, une modification ou une disparition de la passion correspondante devrait en résulter également. Si je désire fortement me donner du plaisir en dégustant cette tarte aux fraises et si j'apprends qu'elle est en caoutchouc ou en plastique, il y a des chances pour que cette information diminue ou élimine mon désir. Mais cette observation triviale n'atteint pas l'argument de l'inertie de la raison, car dans ce cas, elle s'est tout simplement mise au service de la passion correspondante.

Autrement difficile est le cas de nos jugements moraux en matière de justice (*E.P.M.*, p. 246 ; *T.N.M.*, p. 705), et le problème du sentiment moral en général. Là, les pouvoirs de la raison apparaissent immenses. Elle n'est plus vraiment au service d'une passion ; elle la commande, à ce qui semble du moins.

Nos sentiments naturels de bienveillance peuvent nous incliner à accomplir des actes de justice isolés, qui contredisent l'intérêt général. Notre compassion devrait nous interdire d'accepter comme juste une décision qui attribue au riche le fruit du travail

du laborieux en vertu des lois de la propriété, décision que nous pouvons accepter dans une perspective beaucoup plus large. Bref, nos meilleurs sentiments ne sont pas toujours en harmonie avec la justice.

Dans la discussion du respect que nous éprouvons pour la justice, Hume approfondit sa doctrine du sentiment moral. Il essaie de montrer de quelle façon elle peut être comprise par analogie avec sa théorie des sensations.

Si le jugement moral était purement dérivé de nos passions, il ne pourrait jamais nous permettre d'aller au-delà de nous-mêmes ni d'adopter un point de vue impartial, celui que nous prenons lorsque nous approuvons une décision de justice. C'est que, parmi les passions, aucune ne semble être en mesure de nous expliquer le caractère tout à fait spécifique du jugement moral. Nous avons parlé des sentiments de plaisir et de peine : ils peuvent en effet être considérés comme de bons moteurs de l'action. Nous cherchons ce qui nous plaît et nous fuyons ce qui nous déplaît. Et, à la suite de Hume, nous pourrions dire tout simplement que les actions vertueuses sont celles qui nous plaisent et les actions vicieuses, celles qui nous déplaisent. Mais il est bien évident qu'il peut nous arriver de juger vertueuse une action qui ne provoque aucun plaisir immédiat et mauvaise une action qui peut procurer un certain plaisir : c'est ce qui se passe lorsque nous reconnaissons comme vertueuses certaines décisions de justice qui se font au détriment de notre tendance naturelle à aider les plus faibles.

Parmi les passions qui pourraient être au principe du jugement moral, l'amour est sans doute la plus intéressante. Cependant la passion d'amour qui nous donne autrui comme objet s'applique à ce qui est proche, contigu. Elle n'est pas susceptible de s'étendre sans perdre sa valeur, son intensité, à ce qui est trop éloigné dans l'espace et le temps. Hume suggère qu'il existe un sentiment proche de l'amour, la sympathie, sentiment moins vif et par conséquent plus susceptible de s'appliquer, sans s'édulcorer, à des êtres lointains plus dissemblables.

Mais la sympathie naturelle ne suffit pas encore à nous faire comprendre le respect ou l'admiration pour certaines décisions impartiales, qui contribuent au bien public, car nous savons très bien que notre sympathie peut varier sans que varie notre

estime. « Ce sont les mêmes qualités qui nous touchent en Chine et en Angleterre chez les anciens Grecs, si éloignés de nous et de nos amis les plus proches. » (*T.N.H.*, 706.)

Que signifie l'estime ? Deleuze dit : l'intégrale des sympathies¹. Selon Deleuze, Hume considère que le problème moral et social consiste à passer des sympathies réelles qui s'excluent à un « tout réel qui inclut les sympathies. Il s'agit d'étendre la sympathie ». Il n'est pas du tout sûr que Hume soit aussi optimiste. L'estime n'est pas un sentiment dérivé de l'amour ou de la sympathie : il ressemble plutôt à un mouvement froid ou calme qui se traduit dans une construction artificielle telle que la justice, laquelle en retour est accompagnée de sentiment. Le point de vue impartial du jugement moral, tel qu'il est appréhendé dans notre estime ou dans l'approbation des décisions de justice, est au-delà des passions.

Dans un passage étourdissant de *l'Enquête sur les principes de la morale*, Hume se plaît à imaginer la vie morale des habitants de Fourli, pays imaginaire dans lequel un voyageur a l'impression que toutes les actions qui nous indignent sont couvertes d'éloge : inceste, parricide, esclavage, lâcheté, absence de réciprocité, mensonge. Puis il dévoile que c'est un récit à clé : les habitants de Fourli ne sont que nos Brutus, Socrate, Thémistocle, etc. maquillés, les Athéniens et les Romains tant admirés (*E.P.M.*, p. 191). Mais il parvient à montrer, comme un bon anthropologue contemporain, que ces actes sont justifiés par les principes qui nous plaisent (*E.P.M.*, p. 196).

Lévy-Bruhl n'a pas tort de dire que les moralistes du sentiment naturel présupposent la constance de la nature humaine. « Hume répète après Fontenelle, et avec les Encyclopédistes que les hommes d'aujourd'hui sont aussi semblables à ceux d'autrefois que les chênes et les peupliers de nos campagnes le sont à ceux d'il y a cinq mille ans. Veut-on reconnaître le mécanisme, et le jeu des passions chez nos contemporains ? Il suffit d'étudier Démocrène et Tacite. Voltaire ne tient pas un autre langage². »

Lévy-Bruhl ne se prive pas de railler cette doctrine, « Partout l'amour, l'ambition, l'amour-propre, l'avarice, l'envie, et les

1. Deleuze G., *Empirisme et subjectivité*, Paris, P.U.F., 1973, p. 27.

2. Lévy-Bruhl L., *op. cit.*, 1971, p. 71.

autres passions naissent des mêmes causes, traversent les mêmes crises et produisent les mêmes effets. "C'est tout comme icy" selon la formule que Leibniz emprunte à Arlequin¹.»

Mais il serait peut-être trop cruel de rappeler le destin des doctrines de Levy-Bruhl et celui du relativisme radical en général. De la confrontation intellectuelle entre Hume et Levy-Bruhl, c'est sans doute Hume qui s'en est le mieux sorti. Au demeurant, sa justification de l'universalité des jugements moraux, à partir de l'approbation universelle des mêmes actions, et des mêmes caractères, anticipe d'une certaine façon, certaines idées de Durkheim.

Hume estime qu'il existe une qualité commune à l'ensemble des actions et des caractères universellement approuvés, c'est la sociabilité, une qualité qui contient toutes les autres (*E.P.M.*, p. 30). Bref, on pourrait dire, sans abuser, que Hume prépare Durkheim en affirmant que la morale, ce n'est rien d'autre, partout et toujours, qu'un culte rendu à la société. Mais Hume ajoute une note utilitariste que Durkheim n'a probablement pas appréciée. D'après David Hume, ce que nous approuvons naturellement, ce que nous louons universellement, c'est tout ce qui concourt au bien-être de notre communauté, ou mieux encore au bien-être de la société humaine en général: « Somme toute, il semble indéniable que rien ne peut conférer plus de mérite à une créature humaine que le sentiment de bienveillance parvenu à un degré éminent et qu'une partie au moins de son mérite naît de sa tendance à servir les intérêts de notre espèce et à donner du bonheur à la société humaine. » (*E.P.M.*, p. 36.)

Mais, comme Hume le reconnaît volontiers, ces deux aspects, bienveillance, d'une part, et tendance à agir en vue du bien-être de la société, d'autre part, ne coïncident pas nécessairement. La bienveillance ne va jamais assez loin, ou, plus exactement, elle ne va jamais naturellement là où elle devrait aller. Nos meilleurs sentiments ne suffisent pas. Nos réactions les plus vertueuses sont socialement construites. Nos réactions morales authentiques sont des institutions, des artifices.

Pour certains exégètes, il y aurait dans les thèses morales de Hume une difficulté insurmontable, dont l'origine se trouve pré-

1. *Id.*, p. 71.

cisément dans cette définition « artificialiste » des réactions morales authentiques. Mais on pourrait aussi bien dire que Hume n'est pas aussi attaché à la défense des morales du sentiment naturel qu'on le dit habituellement.

Quoi qu'il en soit, il me paraît difficile de tirer des thèses de Hume un bon argument en faveur du compatibilisme par exclusion. Les commentateurs qui ont cru trouver cette thèse dans ses écrits sont probablement allés au-delà de ce que Hume lui-même proposait explicitement. Ils ont plutôt défendu des arguments de leur cru, parfois brillants, comme ceux de Strawson.

Mais ils ont dû, de toute façon, atténuer une certaine tension. Le compatibilisme par exclusion dit que nos réactions morales d'approbation et de réprobation sont quasi-instinctives et qu'elles sont, de ce fait, immunisées à l'égard de nos connaissances. Comme dans le cas de la vision, ce que nous savons ou croyons n'aurait aucune influence sur nos sensations. C'est bien pourquoi le compatibiliste par exclusion estime que les idées que nous nous faisons du déterminisme, qui sont de toute évidence abstraites et générales, n'ont en réalité aucune influence sur nos réactions d'éloge ou de blâme ni sur la perception que nous avons d'autrui et de nous-mêmes comme êtres libres et responsables.

Mais si nos réactions authentiquement morales ne sont que des constructions artificielles ou institutionnelles, comme le dit Hume, pourquoi devraient-elles être immunisées à l'égard de nos croyances générales comme le sont, semble-t-il, nos réactions sensorielles naturelles ?

Si on veut vraiment comprendre pourquoi le point de vue que nous adoptons dans la querelle du déterminisme n'a, en définitive, pratiquement aucune influence sur nos réactions morales et, indirectement, sur nos engagements envers les idées de liberté et de responsabilité, il faut probablement chercher dans une tout autre direction.

3. L'AUTONOMIE DE L'ÉTHIQUE

Les naturalistes s'intéressent beaucoup à la question de savoir si un être humain à qui manquerait telle ou telle partie du cerveau resterait capable, néanmoins, de porter un jugement moral ou d'agir moralement.

Mais lorsqu'on prend conscience de toutes les difficultés que soulève la réflexion morale, on risque plutôt de se demander comment il se fait que, même avec toute sa tête (un cerveau complet, en parfait état de marche, des facultés psychologiques « épanouies »), on peut cependant rester incapable de savoir quoi penser et quoi faire (ou comment faire) dans la plupart des cas qui nous préoccupent véritablement.

Il est possible que certaines de ces difficultés viennent du fait qu'on passe trop de temps désormais à chercher du côté des déficiences physiologiques, psychologiques ou sociales éventuelles des agents dits moraux et pas assez, ou plus assez, du côté de la réflexion rationnelle sur les systèmes éthiques en tant que tels. Pour le dire autrement, et peut-être plus simplement, nous avons peut-être du mal à progresser dans ces questions parce que nous avons du mal à admettre l'autonomie de l'éthique en tant que sujet théorique, alors que c'est une idée qui est largement acceptée pour la logique ou les mathématiques.

En réalité, ce qu'il faut se demander, c'est dans quelle mesure ce que nous apprennent les différentes sciences naturelles spécialisées peut nous aider à clarifier, et peut-être à résoudre, certaines questions qui se posent dans la réflexion éthique.

Ceux qui défendent l'autonomie de l'éthique estiment que la contribution de ces sciences spécialisées à l'élucidation de ces questions est nulle pour des raisons de principe. Telle est du moins l'idée que défend, entre autres, Thomas Nagel, qui pose le problème de la façon suivante : « Si l'éthique n'est qu'un ensemble de comportements ou d'habitudes soutenues par certaines réponses émotionnelles, on devrait s'attendre à ce que les théories biologiques, par exemple, nous en apprennent beaucoup sur ces

sujets. Mais si c'est une enquête théorique qui peut être menée au moyen de méthodes rationnelles, une enquête dont les critères de justification et de critique sont internes, alors toutes les tentatives de la comprendre de l'extérieur pour ainsi dire, au moyen de la biologie par exemple, perdront beaucoup de leur pertinence¹.»

Nagel semble vouloir dire que les critères de critique ou d'évaluation de la justesse d'une théorie ne peuvent pas être empruntés à un autre domaine théorique. Les techniques qui sont utilisées pour prouver qu'un raisonnement logique n'est pas correct sont internes à la logique. On peut expliquer par des causes psychologiques pourquoi on s'est trompé. On peut trouver miraculeux le fait que la nature nous ait doté des capacités de faire des raisonnements justes ou de construire des normes de raisonnement idéales à partir de nos raisonnements naturels. Mais toutes ces données psychologiques sont mises de côté, pour ainsi dire, lorsqu'on discute d'une question logique particulière. Il devrait en aller de même en éthique, si c'est une enquête théorique.

Une explication biologique ou psychologique de la validité d'un raisonnement logique ou mathématique complexe pourrait nous paraître futile. De même, une explication biologique ou psychologique de la supériorité de l'éthique de Kant sur l'éthique utilitariste ou l'inverse, une telle explication pourrait nous sembler extravagante, et en définitive complètement à côté de la question.

On pourrait trouver beaucoup d'autres arguments directs, indirects en faveur de l'autonomie de l'éthique. Voici les trois principaux, à mon sens :

1) l'argument de la distinction entre les causes de nos croyances et les justifications de nos croyances ;

2) l'argument de la question ouverte, c'est-à-dire de l'impossibilité de définir le bien en termes non éthiques (proposé par Moore, au début de ce siècle, et qui a remarquablement survécu à toutes les tentatives de réfutation, une résistance qui n'est peut-être pas à mettre au crédit de l'argument !) ;

3) l'argument de la parcimonie normative (plus ou moins inspiré de Kant).

1. Nagel T., « Ethics as an Autonomous Theoretical Subject », in Stent G., *Morality as Biological Phenomenon*, Berkeley, University of California Press, 1978, p.196-205.

Je ne vais pas développer en détail chacun de ces arguments : je vais seulement donner quelques exemples qui pourront nous permettre d'identifier les problèmes.

Commençons par la distinction entre les *causes* et les *justifications des croyances*. Elle correspond, en gros, à la distinction proposée par Wittgenstein entre les causes et les raisons. C'est une distinction que nous faisons très spontanément, je crois, lorsque les croyances ou les comportements nous semblent scandaleux, répugnants, mauvais ou défectueux et plus difficilement lorsque ces croyances ou ces comportements nous semblent corrects, bons, ou admirables. Ainsi, il semble clair que donner les causes psychologiques ou biologiques de la xénophobie, du racisme, de la haine, de l'intolérance ne revient absolument pas à les justifier rationnellement. Mais en même temps, beaucoup semblent penser que donner les causes biologiques ou psychologiques de l'altruisme ou de la bienveillance revient en réalité à les justifier, en leur donnant un fondement naturel !

C'est une attitude incohérente. Il me semble qu'il faut adopter une attitude plus équitable et reconnaître que les causes psychologiques ou biologiques qui peuvent expliquer nos bons comportements ne peuvent absolument pas tenir lieu de justification à ces comportements, de la même façon que les causes psychologiques ou biologiques qui peuvent expliquer nos comportements répugnants ne peuvent pas du tout justifier ces comportements.

On peut donner une signification beaucoup plus forte encore à l'opposition entre les causes de nos croyances et les justifications de nos croyances, si on estime que nos croyances ne peuvent être justifiées en définitive que si elles correspondent à une réalité indépendante de nos croyances, de nos attitudes et de nos sentiments.

Cette thèse, dite objectiviste implique que le fait que nous ayons évolué comme des organismes capables de comprendre que « $2 + 2 = 4$ » ou qu'il vaut mieux ne pas se faire écraser par un éléphant ne signifie absolument pas que le résultat d'un calcul arithmétique ou les propriétés d'un éléphant soient des produits de nos préférences même raisonnées ou de nos imaginations. Ce sont des réalités autonomes.

C'est une chance, en un certain sens, que nous sachions ce que peut faire un éléphant, mais l'éléphant aurait pu faire ce qu'il fait même si nous n'avions eu aucun moyen de le découvrir.

Tout le monde est disposé à admettre qu'il y a des objets que les mollusques sont incapables de percevoir. Mais personne n'en tire la conclusion que ces objets n'existent pas. De la même façon, il se pourrait qu'il y ait des objets physiques ou mathématiques et même des objets moraux que nous, humains, sommes actuellement ou pour l'éternité, dans l'incapacité de percevoir. Mais cela ne veut pas dire que ces objets n'existent pas.

Cependant, il faudrait aller beaucoup plus loin, on l'a vu, pour rendre intelligible cette thèse dite réaliste lorsqu'elle s'applique à des jugements moraux tels que « L'esclavage est un mal » ou « La tolérance est un bien ». Je préfère la laisser de côté et passer au second argument.

Ce deuxième argument, dit de la *question ouverte*, est celui de Moore¹. Je l'ai évoqué précédemment ; je me contenterai d'en rappeler les grandes lignes.

Supposons que vous ayez affirmé : « La tolérance est un bien ». On peut vous demander : mais qu'entendez-vous par bien ? Si vous acceptez de répondre, vous essaieriez probablement de donner un contenu à votre idée du bien. Vous direz par exemple : bien signifie agréable, ou impartial, ou fonctionnel, ou rationnel, ou commandé par la loi divine, ou utile à la société dans son ensemble, ou propre à satisfaire nos besoins fondamentaux, ou notre bien-être, ou notre bonheur, ou notre survie en tant qu'espèce. Mais on pourra toujours vous demander, pour chacune de ces définitions, en quoi est-ce un bien ? En quoi l'utile est-il bien ? En quoi le rationnel est-il bien ? Pourquoi ce qui est commandé par la loi divine ou ce qui est utile à la société ou propre à satisfaire notre bien-être est-il bien ? En quoi le fait de survivre en tant qu'espèce est-il bien ?

En fait, l'argument de Moore revient à dire qu'en essayant de définir le bien en termes non éthiques, on commet une erreur de raisonnement qu'on retrouve chaque fois qu'on essaie de définir des termes primitifs ou des notions dont on ne peut qu'avoir l'intuition. Cette erreur, Moore l'a appelée le sophisme ou le paralogisme naturaliste. C'est, d'après lui, l'erreur la plus commune et la plus désastreuse dans le traitement des questions morales.

1. Moore G.E., *op. cit.*, 1903, p. 6-22.

Moore estime qu'il faut se garder d'essayer de traduire le terme « bien » ou de le réduire en empruntant un vocabulaire non éthique. C'est en ce sens qu'il défend l'autonomie de l'éthique. Je reconnais que cette défense commence assez bien — et finit plutôt mal, et c'est pourquoi je vais évoquer un troisième argument.

J'ai choisi de l'appeler argument de la *parcimonie normative*, car il dit qu'il est inutile de multiplier les normes sans nécessité. C'est un argument un peu plus traditionnel, qu'on peut retrouver, si on n'est pas trop regardant, dans l'approche kantienne.

Cet argument repose sur le principe qu'il est absurde d'interdire ce que personne ne ferait en aucun cas et inutile de prescrire ce que tout le monde ferait de toute façon. Tout le monde comprend qu'il serait absurde d'interdire à quelqu'un de manger sa propre tête, et Kant avait fait remarquer qu'un commandement ordonnant à chacun de chercher à se rendre heureux serait une sottise car on n'ordonne jamais à quelqu'un ce qu'il veut inévitablement de lui-même¹. Si la moralité peut nous apparaître comme devoir, obligation, impératif, c'est précisément parce qu'elle porte sur ce que nous ne voulons pas naturellement, si on peut dire, ou plus exactement, parce que ce que nous voulons ou ne voulons pas naturellement, parce que nos inclinations (comme le dirait Kant) ne doivent pas rentrer en ligne de compte dans la détermination de ce qui est moral.

Le troisième argument en faveur de l'autonomie de l'éthique est plus indirect, et je ne suis pas sûr qu'il soit exactement du même genre que les deux premiers. Quoi qu'il en soit, il contribue, bien que sur un autre plan, à justifier l'autonomie de l'éthique.

Je vais essayer à présent de montrer comment ces considérations d'autonomie interviennent dans l'examen d'un sujet moral précis. J'ai choisi l'exemple de l'altruisme parce qu'il est au cœur de la plupart des tentatives de réduction sociobiologiste de l'éthique, qui sont en réalité des façons de nier l'autonomie de l'éthique en tant que sujet théorique².

La plupart des gens qui ont réfléchi à la question de l'altruisme ont admis qu'elle posait deux problèmes : un problème de défini-

1. Kant, E., *Critique de la raison pratique* (1788), trad. J. Gibelin, Paris, Vrin, 1974, p. 51.

2. Changeux J.P., éd., *op. cit.*, 1993.

tion et un problème de justification morale. Il est assez courant, aujourd'hui, de dire que les comportements altruistes ne sont pas réservés à l'homme, et qu'on en trouve des traces dans toutes les sociétés animales. Il suffit en effet de proposer une définition minimale, dispositionnelle ou comportementale de l'altruisme pour justifier cette extension. Une définition dispositionnelle ou comportementale de l'altruisme pourrait avoir l'allure suivante : disposition au sacrifice ou acte de sacrifice de ses intérêts propres au bénéfice des intérêts d'autrui et de sa communauté d'appartenance. Or il semble bien que l'oiseau qui pousse un cri pour alerter ses congénères de la présence d'un prédateur se comporte de façon altruiste, puisqu'il sacrifie sa propre vie en attirant le prédateur par son cri, tout en favorisant la survie de ses congénères qui, alertés, peuvent se sauver.

Mais en faisant observer qu'une interprétation égoïste de ce comportement n'était pas exclue, en termes d'intérêt génétique par exemple, certains sociobiologistes ont montré, indirectement, qu'une telle définition était insatisfaisante. Elle l'est encore plus, bien sûr, lorsqu'on essaie de l'appliquer aux comportements humains.

Étant donné qu'il semble y avoir dans les affaires humaines une sorte d'arbitraire du signe social par rapport aux motivations, tout comportement dit altruiste peut recevoir une interprétation égoïste. C'est le cas par exemple, lorsqu'on estime qu'un acte généreux n'est qu'une sorte d'invitation à la réciprocité, ou un investissement social à long terme. En fait, la question est bien celle de la motivation de l'acte ou plus exactement de l'*intention* dans lequel il est accompli.

D'autre part, il importe que l'identification de l'intention autorise une interprétation en termes de liberté ou de responsabilité, c'est du moins ce que j'ai essayé de montrer. Si on n'a absolument aucune raison de dire que l'oiseau a choisi, en un sens important quelconque, de sacrifier sa vie, je crois qu'on aura beaucoup de réticences à dire qu'il *était* altruiste, même s'il s'est *comporté* de façon altruiste. Il en va de même dans les affaires humaines, bien sûr. Mais commençons par la question simple des motifs.

Un comportement ne sera dit altruiste, semble-t-il, que s'il est suffisamment clair qu'il n'a pas été accompli dans l'intention de recevoir, immédiatement ou à plus long terme, un bénéfice person-

nel en termes de prestige ou de pouvoir ou d'argent et, également, si l'intention dans laquelle il a été accompli n'était pas mauvaise en elle-même, comme lorsqu'on peut envisager la possibilité que le comportement n'était qu'une façon de s'immiscer dans les affaires d'autrui, ou de vouloir se prendre pour un saint ou pour Dieu sur terre, avec tous les avantages et les désavantages psychologiques de ce genre de comportement mégalomane.

C'est pour toutes ces raisons que les tentatives d'identifier l'altruisme en termes purement comportementaux peuvent paraître vaines. Mais aussitôt qu'on convient de la nécessité d'introduire l'idée d'intention consciente ou inconsciente dans la description de l'altruisme, on se retrouve dans une situation extraordinairement inconfortable du point de vue scientifique. Il y a des scientifiques et des philosophes qui rejettent entièrement l'idée d'intention. Mais même ceux qui acceptent de l'utiliser reconnaissent ses limites. Il est très difficile, par exemple, d'expliquer comment il est possible d'intégrer l'idée d'intention dans un schéma explicatif causal, c'est-à-dire d'expliquer comment une entité abstraite ou purement psychologique peut entraîner un quelconque mouvement du corps. C'est le très vieux problème de la relation corps-esprit.

La solution proposée par certains physicalistes aujourd'hui consiste à réduire l'intention à un certain état du cerveau. Mais il semble qu'alors toutes propriétés spécifiques de l'intention disparaissent ou que dans le meilleur des cas l'intention n'apparaisse que comme une sorte d'épiphénomène, une sorte de représentation sans aucun pouvoir causal.

Il existe, bien sûr, des tentatives de donner un statut naturaliste non réductionniste aux intentions et aux représentations en général. Mais la question reste ouverte, et je ne crois pas que nous ayons les moyens de la résoudre dans un cadre naturaliste. La raison pour laquelle il me semble qu'on ne peut pas résoudre la question de l'intention dans un cadre naturaliste, c'est qu'elle est étroitement liée à la notion de liberté.

Nous utilisons la notion d'intention dans des contextes très spécifiques. C'est, en général, lorsqu'on soupçonne quelqu'un de s'être mal conduit qu'on se pose généralement la question de l'intention et cela vaut aussi pour les comportements altruistes, lorsqu'on se demande s'ils ne sont pas hypocrites, par exemple.

Et c'est bien parce qu'elle est liée à la notion de liberté que la notion d'intention semble autoriser le jugement moral : on n'accuse pas, comme le disait déjà Aristote, l'aveugle de naissance pour son infirmité. On porte des jugements d'éloge et de blâme si on estime que l'agent était libre, ou que son comportement était intentionnel en un sens important quelconque. Si la notion d'intention était interprétée dans un sens purement déterministe, elle n'aurait pas grand-chose à voir avec celle que nous utilisons habituellement. Et si nous l'utilisons dans un sens non déterministe, il nous faut, semble-t-il, admettre d'une façon ou d'une autre, que nous pouvons identifier un comportement en termes non naturalistes, même si à la manière de Kant, nous admettons, en même temps, qu'une interprétation naturaliste de ce comportement est aussi possible et en un certain sens nécessaire.

On voit donc qu'il y a déjà, dans la simple identification du phénomène, de l'altruisme des éléments moraux. Ils deviennent plus évidents encore, lorsque, acceptant de modérer un peu les prétentions naturalistes, on admet une définition complexe de l'altruisme qui fait intervenir la notion de la liberté de l'intention et qu'on se pose la question de sa valeur morale.

Il est curieux d'observer que la plupart des théories naturalistes qui placent l'altruisme au cœur de leur construction ne se posent pas du tout la question de la valeur morale de l'altruisme, comme s'il allait de soi que l'altruisme était un bien. Or, il y a deux objections traditionnelles à l'idée que l'altruisme est un bien¹.

La première porte sur la contradiction ou le conflit entre la valeur de l'autonomie personnelle et de l'accomplissement de soi d'une part, et la valeur du sacrifice, de l'autre. Les hommes et les femmes qui passent plus de temps à s'occuper d'autrui que d'eux-mêmes, nous paraissent suspects parce qu'on estime, semble-t-il, que le souci de soi est une valeur supérieure et que l'altruisme n'est acceptable que s'il ne détruit pas ce souci.

La deuxième objection ne porte pas sur le sujet du comportement altruiste mais sur son objet. Il n'est pas plus facile de reconnaître les intentions d'autrui que les siennes. Essayer de faire le bien d'autrui peut conduire aux pires catastrophes, c'est bien connu. L'objection anti-altruiste du deuxième genre dit donc que

1. Paul E.F., Miller F.D. Jr, Paul J., édés, *Altruism*, Cambridge, Cambridge University Press, 1993.

le comportement altruiste doit être subordonné au souci de l'autonomie de son objet.

En règle générale, les théories morales ne semblent pas donner à l'altruisme une valeur intrinsèque. L'altruisme ne vaut que s'il peut être considéré comme un comportement rationnel, ce qui est loin d'être évident, ou comme un comportement susceptible de favoriser l'épanouissement personnel et celui d'autrui.

Mais on n'est pas encore sorti d'affaire lorsqu'on a subordonné l'altruisme à la rationalité ou au bonheur, car la question de savoir si les objectifs moraux sont indiscutables est loin d'être tranchée.

Les sociobiologistes et les philosophes qui acceptent quelques-unes, au moins, de leurs thèses semblent donner au bonheur ou au bien-être ou à la survie de l'espèce une valeur suprême.

Il est difficile de trouver un fondement non théologique à la valeur morale de la survie de l'espèce en tant que telle. Quant à l'idée de bien-être ou de bonheur en tant qu'objectif de la visée éthique, elle est loin d'être irrécusable, on le sait bien.

À côté de toutes les critiques kantiennees, on peut estimer, tout simplement, que beaucoup de choses nous importent éminemment qui sont loin de nous apporter le bonheur. Nous pouvons avoir beaucoup de raisons de préférer une vie de risque, d'aventure, de création, d'angoisse et même de souffrance, à la vie paisible du bœuf ou du sage.

BIBLIOGRAPHIE

- Abell P., «Homo Sociologicus : Do We Need Him/Her ? », *Sociological Theory*, 9, 2, 1991, p. 195-198.
- Alexander J., « The Parsons Revival in Germany », *Sociological Theory*, 2, 1984, p. 394-412.
- Anscombe G.E.M., *Intention*, Oxford, Basil Blackwell, 1957.
- Anscombe G.E.M., «The First Person», in Guttenplan S., éd., *Mind and Language*, Oxford, Clarendon Press, 1975, p. 45-65.
- Anscombe G.E.M., «Modern Moral Philosophy», dans *Ethics, Religion and Politics, Collected Philosophical Papers*, III, Oxford, Basil Blackwell, 1981, p. 32.
- Aristote, *L'Éthique à Nicomaque*, trad. J. Tricot, Paris, Vrin, 1983.
- Aristote, *De la génération et de la corruption*, éd. et trad. C. Mugler, Paris, Les Belles Lettres, 1966.
- Aristote, «De l'amitié et de la haine», *Rhétorique* II, éd. et trad. M. Dufour, Paris, Les Belles Lettres, 1967, p. 68-72.
- Ashby R.W., «Linguistic Theory of the A Priori », dans *The Encyclopedia of Philosophy*, Edwards P., éd., New York, Free Press, MacMillan, 1967.
- Austin J.L., «Performatif-Constatif», dans *La philosophie analytique*, Paris, Minuit, 1962, p. 271-304.
- Austin J.L., *Quand dire, c'est faire* (1962), trad. G. Lane, Paris, Le Seuil, 1970.
- Austin J.L., *Écrits philosophiques* (1970), trad. L. Aubert et A.L. Hacker, Paris, Le Seuil, 1994.
- Austin J.L., *Philosophical Papers*, 2^e éd., Oxford, Oxford University Press, 1970.
- Audi R., «Mental Causation : Sustaining and Dynamic» dans Heil J. et Mele A., eds, *Mental Causation*, Oxford, Clarendon Press, 1993, p. 53-74.
- Baker G.P., Hacker P.M.S., *Wittgenstein : Rules, Grammar and Necessity*, Oxford, Basil Blackwell, 1985.
- Bicchieri C., «Norms of Cooperation», *Ethics*, 100, 1990, p. 838-861.
- Bierstedt R., *The Social Order*, 2^e éd., New York, Mac Graw-Hill Book Co, 1963.
- Bergson H., *Le Rire*, (1924), Paris, P.U.F., Quadrige.
- Blackburn S., «Supervenience Revisited» dans Hacking I., *Exercises in Analysis, Essays by Students of Casimir Levy*, Cambridge, Cambridge University Press, 1985, p. 47-67.
- Blackburn S., *Essays in Quasi-Realism*, Oxford, Oxford University Press, 1993.

- Boatright J.R., «Hume's Account of Moral Sentiment», *Revue internationale de philosophie*, 115-116, 1976, p. 79-90.
- Boghossian P.A., «Analyticity», dans Wright C., Hale B., *A Companion to the Philosophy of Language*, Oxford, Basil Blackwell, 1995.
- BonJour L., *The Structure of Empirical Knowledge*, Cambridge, Mass. Harvard University Press, 1985.
- Boudon R., *Le Juste et le Vrai. Études sur l'objectivité des valeurs et de la connaissance*, Paris, Fayard, 1995.
- Bourdieu P., *Ce que parler veut dire*, Paris, Fayard, 1982.
- Bouveresse J., *Wittgenstein : la rime et la raison. Science, éthique et esthétique*, Paris, Minuit, 1973.
- Bouveresse J., *Le mythe de l'intériorité. Expérience, signification et langage privé chez Wittgenstein*, Paris, Minuit, 1976 (2^e éd. avec une nouvelle préface, 1987).
- Bouveresse J., «Frege, Wittgenstein, Dummett et la nouvelle "querelle du réalisme" », *Critique*, 399-400, 1980, p. 881-896.
- Bouveresse J., *La force de la règle. Wittgenstein et l'invention de la nécessité*, Paris, Minuit, 1987.
- Bouveresse J., *Le pays des possibles. Wittgenstein, les mathématiques et le monde réel*, Paris, Minuit, 1988.
- Bouveresse J., *Philosophie, mythologie et pseudo-science, Wittgenstein, lecteur de Freud*, Combas, L'éclat, 1991.
- Brink D.O., *Moral Realism and the Foundation of Ethics*, Cambridge, Cambridge University Press, 1989.
- Bubner R., «Norme et histoire», *Critique*, 413, 1981, p. 929-949.
- Bunge M., «The Revival of Causality», dans Floistad G., éd., *Contemporary Philosophy* (2), La Haye, M. Nijhoff, 1982, p. 133-135.
- Canto-Sperber M., *La philosophie morale britannique*, Paris, P.U.F., 1994.
- Cayla F., *Routes et déroutes de l'intentionnalité. La correspondance R. Chisholm-W. Sellars*, Combas, L'éclat, 1991.
- Changeux J.P., *Fondements naturels de l'éthique*, Paris, Odile Jacob, 1991.
- Chazel F., «Norme sociale», *Encyclopédie universelle de philosophie*, Paris, P.U.F. 1991, p. 1768-1769.
- Chisholm R.M., *Person and Object*, Allen & Unwin, Londres, 1976.
- Chisholm R.M., «The Truths of Reason» (1977) dans Moser P.K., *op.cit.*, 1987, p. 122-144.
- Church J., «L'émotion et l'intériorisation des actions», trad. Yves Bellanger, *Raisons pratiques*, 6, (*La couleur des pensées*), Paris, E.H.E.S.S., 1995.
- Cicourel A.V., *La sociologie cognitive* (1972), trad. J. et M. Olson, Paris, P.U.F., 1979.

- Dancy J., «Transatlantic Variants», *T.L.S.*, 23-24, Mars 1990, p. 326.
- Darwall S., Gibbard A., Railton P., «Toward Fin de Siècle Ethics : Some trends», *The Philosophical Review*, 101, 1, 1992, p. 115-189.
- Davidson D., *Paradoxes de l'irrationalité*, trad. P. Engel, Combas, L'éclat, 1991.
- Davidson D., *Actions et événements* (1982), trad. P. Engel, Paris, P.U.F., 1993.
- Davidson D., «La théorie cognitive de l'orgueil de Hume» (1976), dans *op.cit.*, 1993, p. 93-117.
- Davidson D., «Liberté d'action» (1974), dans *op. cit.*, 1993, p. 94-117.
- Deleuze G. *Empirisme et subjectivité*, Paris, P.U.F., 1973.
- Dennett D., *La stratégie de l'interprète* (1987), trad. P. Engel, Paris, Gallimard, 1990.
- Descombes V., «Naturaliser les sciences morales ?», *La Quinzaine Littéraire*, 644, avril 1994.
- Descombes V., *La denrée mentale*, Paris, Minuit, 1995.
- Devitt M., «Dummett's Anti-realism», *The Journal of Philosophy*, LXXX, 1983, p. 73-99.
- Dummett M., *Truth and Other Enigmas*, Londres, Duckworth, 1978.
- Dummett M., *The Interpretation of Frege's Philosophy*, Londres, Duckworth, 1981.
- Dummett M., *The Logical Basis of Metaphysics*, Cambridge, Mass., Harvard University Press, 1991.
- Durkheim E., «Jugements de valeur et jugements de réalité» (1911), *Sociologie et philosophie*, éd. par C. Bouglé, Paris, P.U.F., 1974, p. 102-121.
- Dworkin R., *A Matter of Principle*, Cambridge, Mass., Cambridge University Press, 1985.
- Elster J., *Le laboureur et ses enfants* (1983-1984), trad. A. Gerschenfeld, Paris, Minuit, 1986.
- Elster J., «Norms of Revenge», *Ethics*, 100, 1990, p. 862-885.
- Elster J., *Nuts and Bolts for the Social Sciences*, Cambridge, Cambridge University Press, 1989.
- Engel P., «La quasi-réalité des valeurs et des normes», *Droits*, 18, 1993, p. 99-107.
- Engel P., «Trois formes de normativité», dans Bilgrami A., Davidson D., Engel P., Laurier D., Seymour M., *Lire Davidson. Interprétation et holisme*, Combas, L'éclat, 1994 a, p. 205-225.
- Engel P., *Introduction à la philosophie de l'esprit*, Paris, La Découverte, 1994 b.
- Engel P., «La philosophie de l'esprit et les sciences cognitives», dans *La philosophie anglo-saxonne*, Meyer M., éd., Paris, P.U.F., 1994 c, p. 529-583.

- Evans G., *The Varieties of Reference*, Oxford, Oxford University Press, 1982.
- Everson S., «L'explication aristotélicienne du hasard», *Revue de philosophie ancienne*, 1, 1988, p. 39-76.
- Fauconnier G., «Comment contrôler la vérité : remarques illustrées par des assertions dangereuses et pernicieuses en tout genre», *Actes de la recherche en sciences sociales*, 25, 1979, p. 4-22.
- Field H., «Realism and Relativism», *The Journal of Philosophy*, 1982, p. 553-567.
- Flanagan O., «Quinean Ethics», *Ethics*, 93, 1982, p. 56-74.
- Fodor J.A., *Psychosemantics*, Cambridge, Mass., M.I.T. Press, 1987.
- Fodor J.A., «Everything Else Being Equal; Hedged Laws and Psychological Explanation», *Mind*, 1, 1991, p. 21-34.
- Frankfurt H., «La liberté de la volonté et la notion de personne» (1971), dans Neuberger M., éd., *Théorie de l'action*, Bruxelles, Mardaga, 1991, p. 253-269.
- Frankfurt H., «Le problème de l'action» (1978), dans Neuberger M., *op.cit.*, 1991, p. 241-251.
- Gibbard A., «Norms, Discussion and Ritual : Evolutionary Puzzles», *Ethics*, 100, 1990, p. 787-802.
- Gibbs J.P., «Norms : The Problem of Definition and Classification», *The American Journal of Sociology*, 70, 1965, p. 586-593.
- Goldman A.I., *A Theory of Human Action*, Englewood Cliffs, Prentice Hall, 1970.
- Goldman A.I., «The Volitional Theory Revisited» dans Brand M. et Walton D., éd., *Action Theory*, Reidel, Dordrecht, 1976, p. 67-84.
- Guillaumin J., «Honte, culpabilité et dépression», *Revue française de psychanalyse*, 5-6, XXXVII, 1973.
- Hare R.M., *The Language of Morals*, Oxford, Clarendon Press, 1952.
- Harman G., «Moral Explanations of Natural Facts - Can Moral Claims be Tested Against Moral Reality ?», *The Southern Journal of Philosophy*, Supplément, 1986, p. 57-78.
- Hart H.L.A., «The Ascription of Responsibility and Rights», *Proceedings of the Aristotelian Society*, 49, 1948, p. 171-194.
- Hobbes T., *De la nature humaine* (1651), trad. d'Holbach, Paris, Vrin, 1971.
- Homans G., *Social Behavior : Its Elementary Forms*, New York, Harcourt, Brace and Co, 1961, p. 46.
- Horwich P., «Three Forms of Realism», *Synthese*, 51, 1982, p. 181-201.
- Horwich P., *Truth*, Oxford, Basil Blackwell, 1990.
- Hospers J., *An Introduction to Philosophical Analysis*, Londres, Routledge & Kegan Paul, 1990, p. 103-155.

- Hume D., *Traité de la nature humaine*, 1739-1740, trad. A. Leroy, Paris, Aubier, 1946.
- Hume D., *Les passions* (Livre II du *Traité de la nature humaine*), trad. J.P. Cléro, Paris, Garnier-Flammarion.
- Hume D., *Enquête sur l'entendement humain* (1748), trad. M. Beyssade, Garnier-Flammarion, 1983.
- Hume D., *Enquête sur les principes de la morale* (1751), trad. A. Leroy, Paris, Aubier, 1947.
- Hurley S., *Natural Reasons*, Oxford, Oxford University Press, 1989.
- Isambert F.A., *Rite et efficacité symbolique*, Paris, Le Cerf, 1979.
- Isambert, F.A., «Durkheim et la sociologie des normes» dans Chazel F. et Commaille J., eds., *Normes juridiques et régulation sociale*, Paris, Librairie générale de Droit et de Jurisprudence, 1991, p.51-64.
- Jacob P., «Patterns, causalité mentale et lois intentionnelles», *Les études philosophiques*, 3, 1992, p. 391-413.
- James W., *Précis de Psychologie*, 1908, trad. E. Baudin et G. Bertier, Paris, Rivière, 1929.
- Johnson H.M., *Sociology*, New York, Harcourt Brace and Co, 1960.
- Johnston P., *Wittgenstein and Moral Philosophy*, Londres, Routledge & Kegan Paul, 1989.
- Kant E., *Critique de la raison pure* (1787), trad. A. Tremesaygues et B. Pacaud, Paris, P.U.F., 1975.
- Kant E., *Critique de la raison pratique* (1788), trad. J. Gibelin, Paris, Vrin, 1974.
- Kenny A., *Action, Emotion and Will*, Londres, Routledge & Kegan Paul, 1963.
- Kim J., «Non Causal Connections», *Noûs*, 8, 1974, p. 41-52.
- Kim J., «Causation, Nomic, Subsumption and the Concept of Event», *The Journal of Philosophy*, 8, LXX, 1976, p. 217-236.
- Kolakowski, L., «Normes qui commandent et normes qui décrivent», dans *Normes et déviations*, Neuchâtel, La Baconnière, 1987, p. 15-47.
- Kripke S.A., «A Priori Knowledge, Necessity and Contingency», (1980), p. 145-160, (extrait de *Naming and Necessity*, 1972, Cambridge, Mass., Cambridge University Press, Trad. P. Jacob, P. Recanati, *La logique des noms propres*, Paris, Minuit, 1982.
- Ladrière P., «Le conflit entre rationalité cognitive-instrumentale et rationalité morale-pratique dans la sociologie du droit de Max Weber», dans Chazel F. et Commaille J., eds., *Normes juridiques et régulations sociales*, Paris, Librairie générale de Droit et de Jurisprudence, 1991, p. 265-271.
- Leiris M., *À cor et à cri*, Paris, Gallimard, 1988.
- Levy-Bruhl L., *La morale et la science des mœurs*, Paris, P.U.F., 1971.

- Lewis C.I., «A Pragmatic Conception of the A Priori» (1923) dans Moser P.K., *op.cit.*, 1987, p. 15-25.
- Lewis D., *Convention*, Oxford, Basil Blackwell, 1969.
- Lewis H.D., «L'esprit et le corps selon O'Shaughnessy», *Archives de Philosophie*, 49, 1986, p. 75-97.
- Livet P., «Normes et faits», *Encyclopédie universelle de philosophie*, Paris, P.U.F., 1991, p. 124-132.
- Mc Dowell J., «Values and Secondary Qualities» dans Sayre-McCord G., éd., *Essays on Moral Realism*, Ithaca, Cornell University Press, 1988, p. 106-181.
- Mackie J.L., *Ethics, Inventing Right and Wrong*, Penguin, Books, 1977.
- Melden A.I., *Free Action*, Londres, Routledge & Kegan Paul, 1961.
- Moore G.E., *Principia Ethica*, Cambridge, Cambridge University Press, 1903.
- Morris R.T., «A Typology of Norms», *American Sociological Review*, XXI, 1956, p. 6-10.
- Moser P.K., «A Priori Knowledge», Oxford, Oxford University Press, 1987.
- Nagel T., «Ethics as an Autonomous Theoretical Subject», dans Stent G., *Morality as Biological Phenomenon*, Berkeley, University of California Press, 1978, p. 196-205.
- Nagel T., «La fortune morale», dans *Questions mortelles* (1979), trad. P. Engel et C. Engel, Paris, P.U.F., 1983, p. 38-53.
- Nagel T., *Le point de vue de nulle part* (1986), trad. S. Kronlund, Combas, L'éclat, 1993.
- Neuberg M., Introduction à *Théorie de l'action. Textes majeurs de la philosophie analytique de l'action*, Neuberg M., éd., Bruxelles, Mardaga, 1991.
- Nisbett R. et Wilson T., «On Telling More than We Can Know: Verbal Report on Mental Processes», *Psychological Review*, 84, 1977, p. 231-259.
- Nozick R., *The Nature of Rationality*, Princeton, Princeton University Press, 1993.
- Ogien R., *Un portrait logique et moral de la haine*, Combas, L'éclat, 1993.
- Oldenqvist A., «Choosing, Deciding and Doing», *The Encyclopedia of Philosophy*, Edwards P., éd., New York, The Free Press, McMillan, 1967.
- O'Shaughnessy B., *The Will. A Dual Aspect Theory*, Tomes 1 et 2, Cambridge, Cambridge University Press, 1980.
- Pacherie E., *Naturaliser l'intentionnalité*, Paris, P.U.F., 1993.
- Parsons T., «The Role of Ideas in Social Action» (1938) dans *Essays in Sociological Theory*, Free Press Paperback, New York, 1954, p. 19-33.

- Petit J.L., «La Théorie de l'action : comment la faire ?» dans *La sémantique de l'action*, Ricœur P., éd., Éd. du C.N.R.S., 1977, p. 189-222.
- Petit J.L., *L'action dans la philosophie analytique*, Paris, P.U.F., 1991.
- Pettit P., «Virtus Normativa : Rational Choice Perspectives», *Ethics*, 100, 1990, p. 725-755.
- Pettit P., «Realism and Response-Dependence», *Mind*, 4, 1991, p. 586-626.
- Pharo P., *Le sens de l'action et la compréhension d'autrui*, Paris, L'Harmattan, 1993.
- Platon, *Lois*, I, éd. et trad. E. Des Cases, Paris, Les Belles Lettres, 1951.
- Platts M., «Moral Reality» dans Sayre-McCord G., éd., *Essays on Moral Realism*, Ithaca, Cornell University Press, 1988, p. 282-300
- Plutarque, «De l'Envie et de la Haine», *Œuvres Morales*, VII, 2, éd. et trad. R. Klaerr et Y. Vernière, Paris, Les Belles Lettres, 1974.
- Proust J., *Questions de forme : logique et proposition analytique de Kant à Carnap*, Paris, Fayard, 1980.
- Putman H., «Analyticity and Apriority : Beyond Wittgenstein and Quine» (1979) dans Moser P.K., *op.cit.*, 1987, p. 85-111.
- Putman H., *Raison, vérité et histoire* (1981), trad. A. Gerschenfeld, Paris, Minuit, 1984.
- Putman H., *Le réalisme à visage humain* (1990), trad. C. Tiercelin, Paris, Le Seuil, 1994.
- Putman H., *The Many Faces of Realism*, La Salle, Illinois, Open Court, 1987;
- Quéré L. et Rallet A., Présentation du dossier «Conventions», *Réseaux*, C.N.E.T., 1993.
- Quine W.V.O., «Two Dogmas of Empiricism», dans *From a Logical Point of View*, Cambridge, Mass., Harvard University Press, 1953, p. 20-46, trad. P. Jacob, «Les deux dogmes de l'empirisme», dans Jacob P. éd., *De Vienne à Cambridge*, Paris, Gallimard, 1980, p. 87-113.
- Quine W.V.O., «On The Nature of Moral Values», *Theories and Things*, Cambridge, Mass., Harvard University Press, 1981, p. 56-66.
- Quinn W.S., «Truth and Explanation in Ethics», *Ethics*, 1986, p. 524-544.
- Recanati F., Postface à Austin J.L., *Quand dire, c'est faire*, Paris, Points-Seuil.
- Recanati F., «Communication et cognition. Dan Sperber et Deirdre Wilson : La pertinence, communication et cognition», *L'âge de la science*, 4, (*Philosophie de la logique et philosophie du langage*, I), Paris, O. Jacob, 1991, p. 231-249.
- Recanati F., «Paul Grice et la philosophie du langage ordinaire», *L'âge de la science*, 5, (*philosophie de la logique et philosophie du langage*, II), Paris, O. Jacob, 1993, p. 17-22.

- Ricœur P., *Temps et récit*, I, Paris, Le Seuil, 1983.
- Rocher G., *L'action sociale*, Paris, Points-Seuil, 1968.
- Rorty A.O., éd., *Explaining Emotions*, Berkeley, University of California Press, 1980.
- Russell B., «On the Notion of Cause» (1917) dans *Mysticism and Logic*, Londres, Unwin Paperbacks, 1986, p. 173-199.
- Russell B., *The Analysis of Mind* (1921), Londres, Unwin Paperbacks, 1989.
- Russell B., «Notes on Philosophy», *Philosophy*, 35, 1960.
- Ryle G., *La notion d'esprit* (1949), trad. S. Stern-Gillet, Paris, Payot, 1978.
- Scheler M., *La pudeur*, Paris, trad. M. Dupuy, Aubier, 1952.
- Schiffer S., «Ceteris Paribus Laws», *Mind*, 1, 1991, p. 1-17.
- Schmitz F., *Wittgenstein, la philosophie et les mathématiques*, Paris, P.U.F., 1988.
- Searle J., «L'intentionnalité de l'action et de l'intention», trad. J. Proust, *Critique*, octobre 1980, 401, p. 990-1010.
- Searle J., *L'intentionnalité : essai de philosophie des états mentaux* (1983), trad. C. Pichevin, Paris, Minuit, 1985.
- Searle J., *La redécouverte de l'esprit* (1992), trad. C. Tiercelin, Paris, Gallimard, 1995.
- Solomon R.C., *The Passions*, Notre Dame, University of Notre Dame Press, 1983.
- Sperber D., *Le savoir des anthropologues*, Paris, Hermann, 1982.
- Stendhal, *De l'amour*, Paris, Garnier-Flammarion.
- Strauss L., *Droit naturel et histoire* (1954), trad. M. Nathan et E. de Dampierre, Paris, Champs-Flammarion.
- Strawson P.F., «Freedom and Resentment» dans *Freedom and Resentment and Other Essays*, Londres, Methuen, 1974, p. 1-25.
- Stroud B., «Wittgenstein and Logical Necessity», (1965), dans Moser P.K., *op.cit.*, 1987, p. 68-84.
- Tappolet C., «Les émotions et les concepts axiologiques», *Raisons Pratiques*, (La couleur des pensées), Paris, E.H.E.S.S., 1995.
- Taylor C., *The Explanation of Behavior*, Londres, Routledge & Kegan Paul, 1964.
- Taylor C., «What is Human Agency» dans *Philosophical Papers*, I, *Human Agency and Language*, Cambridge, Cambridge University Press, 1985, p. 17-49.
- Vesey G.N.A., «Volition», *Philosophy*, 36, 138, 1961, p.352-365.
- Virvidakis S., «Variétés du réalisme en philosophie morale», *Philosophie*, 22, 1989, p. 1-35.

- Von Wright G.H., *Norm and Action*, Londres, Routledge & Kegan Paul, 1963.
- Waterlow S., *Nature, Change and Agency in Aristotle's Physics*, Oxford, Clarendon Press, 1982.
- Williams B. *Morality, An Introduction to Ethics*, Cambridge, Cambridge University Press, 1972.
- Williams B., *L'éthique et les limites de la philosophie* (1985), trad. M.A. Lescourret, Paris, Gallimard, 1990.
- Winch P., *The Idea of a Social Science*, Londres, Routledge & Kegan Paul, 1958.
- Wittgenstein L., *Le Cahier bleu et le Cahier brun* (1958), trad. G. Durand, Paris, Gallimard, 1965.
- Wittgenstein L., *Remarques sur les fondements des mathématiques* (1956), trad. M.A. Lescourret, Paris, Gallimard, 1983.
- Wrong D.H., «The Oversocialized Conception of Man in Modern Sociology», *American Sociological Review*, 2, 26, 1961, p. 183-193.

REMERCIEMENTS

La plupart des arguments présentés dans ce livre ont été exposés lors de différents séminaires et journées d'études. Certains ont été évoqués dans des publications (« À la recherche de vérités éthiques », *Revue internationale de philosophie*, 4, 1992, p. 458-486; « Le statut moral des concepts mentaux », *Io, Revue internationale de psychanalyse*, 3, 1993, p. 161-178; « L'objectivité morale », *Critique*, 567-568, 1994, p. 625-641). Je n'ai pas pu m'empêcher de faire allusion à mes soucis « analytiques », si je puis dire, dans des conversations ou des correspondances privées. En définitive, mes arguments se sont sensiblement modifiés et je tiens à remercier celles et ceux qui ont fait tout ce qui était possible pour m'aider à les améliorer: Luc Boltanski, Monique Canto-Sperber, Fabien Cayla, Jean-Pierre Cometti, Vincent Descombes, Pascal Engel, Richard Glauser, François Isambert, Henri-Pierre Jeudy, Jacques Katuszewski, Paul Ladrière, Charles Larmore, Pierre Livet, Kevin Mulligan, Thomas Nagel, Marc Neuberger, Simone Novaes, Albert Ogien, Patricia Paperman, Jean-Luc Petit, Patrick Pharo, Louis Quéré, Stephen Schiffer, Christine Tappolet, Claudine Tiercelin, Daniel Vidal.

L'essai sur « la haine » a été publié dans la revue *Raisons Pratiques*, 6, Paris, E.H.E.S.S., 1995, sous une forme à peine différente. Merci aux responsables de cette publication, qui m'ont autorisé à la reproduire.

Table des matières

Avant-propos	5
Introduction : Ce qu'on ne peut pas « naturaliser »	11
Première partie : Les causes et les raisons	33
1. Prédications et platitudes	35
2. Satisfaction et réflexion	52
3. Trois aspects de l'efficacité rituelle	78
4. Normes, règles, lois, etc.	96
Deuxième partie : La vie intérieure	213
1. Ni mentaliste, ni behavioriste	215
2. La honte	125
3. La haine	141
Troisième partie: L'autonomie de l'éthique	157
1. Réalisme ou relativisme : Quoi de neuf ?	159
2. L'immunité des réactions morales	194
3. L'autonomie de l'éthique	218
Bibliographie	227

Achevé d'imprimer
le 3 novembre 1995
sur les presses de
l'Imprimerie A. Robert
116, bd de la Pomme
13011 Marseille

Dépôt légal : novembre 1995

Les explications par les *raisons* sont-elles réductibles à des explications *causales* ? À première vue, nous avons tout intérêt à nous débarrasser des notions intentionnelles (croyances, désirs, motifs, et autres), dans l'analyse de l'action humaine. Pourtant les behavioristes et les physicalistes les plus optimistes s'aperçoivent bien vite, pour toutes sortes de raisons notamment *morales*, qu'il est impossible de les éliminer. C'est à un examen de certaines conséquences de cette aporie qu'est consacré ce livre. Mais il est aussi une invitation à un dialogue renouvelé entre la philosophie analytique et les sciences humaines. Souvent accusée de prôner la réduction des sciences de l'esprit aux sciences de la nature, c'est en effet la philosophie analytique qui présente aujourd'hui les arguments les plus intéressants en faveur de l'*autonomie* des sciences humaines et de certaines stratégies d'interprétation.

Ruwen Ogien est chargé de recherche au CNRS. Il a publié : *Théories ordinaires de la pauvreté* (1988), *Un portrait logique et moral de la haine* (1989), *La faiblesse de la volonté* (1993), et plusieurs articles consacrés à des questions de philosophie morale et de philosophie de l'action. Il prépare un recueil d'essais sur *Le réalisme moral*, à paraître aux PUF.

ART 87711-136.9

Prix : 148 F

ISBN : 2-87711-136-9

Dépôt légal : novembre 1995

Diffusion Harmonia Mundi

