

NATURE SANS RAISON N'EST QUE RUINE DE LA CONSCIENCE

Par

Daniel Laurier

Moi je dis que les bonbons
Valent mieux que la raison
(chanson traditionnelle)

Résumé

Après une brève analyse lexicographique du mot "conscience" et de certains de ses dérivés, j'isole une notion de "conscience" qui a des liens étroits avec le phénomène de l'intentionnalité et je soutiens qu'il ne saurait y avoir d'explication naturaliste d'aucune forme de conscience qui fasse l'économie d'une explication naturaliste de la notion de contenu intentionnel. J'examine ensuite l'idée que nul ne peut avoir de pensées à moins d'avoir la capacité d'avoir des pensées d'ordre supérieur, ce qui m'amène à distinguer deux types ou niveaux de conscience intentionnelle: la conscience individualiste (ou animale) et la conscience anti-individualiste. Après avoir suggéré que la principale caractéristique du naturalisme philosophique est de nier l'existence de normes non naturelles, je distingue une forme modérée de naturalisme qui admet l'existence de normes naturelles et une forme radicale qui la nie. Dans l'hypothèse où les phénomènes intentionnels seraient irréductiblement normatifs, leur existence soulève donc un problème pour le naturalisme. Je conjecture qu'il y a une explication naturaliste de la conscience si et seulement s'il y a une explication naturaliste des normes de rationalité. Je finis par me rallier à la thèse davidsonienne selon laquelle la conscience individualiste dépend de la conscience anti-individualiste, plutôt que le contraire, et par conclure que la seule forme viable de naturalisme est radicale.

1 Conscience et intentionnalité

On dit volontiers de certains systèmes qu'ils ont conscience (ou sont conscients) de certaines choses ou de certains états de choses, ou qu'ils ont conscience (ou sont conscients du fait) que certaines choses sont le cas. De tels systèmes possèdent ce qu'on peut appeler une conscience intentionnelle ou transitive. On dit aussi de certains phénomènes mentaux, tels que des états intentionnels (croyances, désirs, etc.), des épisodes de pensée ou des expériences, qu'ils sont conscients ou inconscients. On parle parfois dans ce cas de conscience intransitive (ou state consciousness). On dit enfin de certains phénomènes mentaux conscients (en particulier les expériences) qu'ils possèdent des qualités subjectives, que seul celui qui en est le siège peut "éprouver" (c'est-à-dire, dont celui qui en est le siège

peut seul faire "l'expérience"). Il est plausible, *prima facie*, de suggérer que ces qualités subjectives ne sont rien d'autre que des qualités conscientes, au sens de la conscience intransitive, et qu'un phénomène mental conscient (ou une qualité consciente) n'est rien d'autre qu'un phénomène mental (ou une qualité d'un phénomène mental) dont le sujet qui en est le siège a conscience, au sens de la conscience intentionnelle. Si ces suggestions sont correctes, il s'ensuit que toute analyse, explication ou théorie concernant soit la "conscience phénoménale" (c'est-à-dire, la conscience des qualités subjectives de l'expérience), soit la conscience intransitive, dépend d'une analyse, explication ou théorie de la conscience intentionnelle, mais qu'elles soient ultimement correctes ou non, elles présentent l'immense avantage de promettre une certaine unification du domaine particulièrement cahotique des "consciousness studies".

La première question à se poser est donc celle de savoir en quoi consiste le fait, pour un système, d'avoir conscience de ou que quelque chose, et s'il y a une manière d'expliquer cette notion qui permette de valider les conjectures du paragraphe précédent. La première chose à remarquer, c'est qu'un énoncé comme

(1) S a conscience (est conscient) de l'objet x

est analogue à un énoncé comme

(2) S connaît l'objet x,

en ce qu'il implique l'existence de l'objet x, et qu'un énoncé comme

(3) S a conscience (est conscient du fait) que p

est analogue à un énoncé comme

(4) S sait que p,

en ce qu'il implique que p.

Je ne veux pas suggérer, en faisant ces rapprochements, que la conscience intentionnelle est une forme de savoir ou de connaissance, car il me semble aller de soi (i) que si on ne peut savoir que p sans être justifié de croire que p, on peut par contre avoir conscience que p sans être justifié de croire que p, (ii) qu'on peut connaître un objet même au moment où on n'en a pas conscience et (iii) qu'on peut avoir conscience d'un objet sans en savoir grand chose. Mais ce que ces analogies permettent de mettre en évidence, c'est que la conscience intentionnelle se rapproche de la croyance vraie, et qu'aucune conception purement internaliste de la conscience (intentionnelle) ne peut par conséquent être correcte. En d'autres termes, de deux systèmes identiques en tous points sauf pour certaines de leurs propriétés relationnelles (c'est-à-dire, sauf pour le fait que x existe ou que p est vrai dans le monde ou l'environnement de l'un mais pas dans celui de l'autre) l'un peut avoir conscience de x ou que p, et l'autre pas. Cela suggère, plus précisément, que le fait d'avoir conscience de x ou que p ne consiste pas exclusivement dans le fait d'être dans un certain type d'état

interne, mais à tout le moins, dans le fait d'être dans un certain type d'état interne qui entretient une certaine relation causale avec l'objet x ou le fait que p.

Il est à noter que rien de ce qui précède n'implique une conception externaliste du contenu intentionnel. Certes, S ne peut avoir conscience de x ou que p sans être dans un état interne qui "représente" l'objet x ou le fait que p; et j'ai suggéré que S a conscience de x ou que p seulement s'il est dans un état interne qui entretient une certaine relation causale avec l'objet x ou le fait que p (ce qui suppose que x existe et qu'il est le cas que p), mais il ne s'ensuit pas que S est dans un état interne qui "représente" l'objet x ou le fait que p seulement s'il est dans un état interne qui entretient une certaine relation causale avec l'objet x ou le fait que p.

Affirmer que S a conscience de x ou que p, ce n'est sans doute pas la même chose qu'affirmer que S a une représentation de x qui est causée par (l'existence de) x ou que S a une représentation du fait que p qui est causée par le fait que p; mais il n'en reste pas moins que la conscience intentionnelle semble bien être un genre d'attitude doxastique, et que le fait que S pense ou croit que p semble être au moins une condition nécessaire pour que S ait conscience que p. Qu'est-ce qui distingue celui qui pense ou croit que p de celui qui a conscience que p, lorsqu'il est le cas que p, et que la croyance ou la conscience que p est causée par le fait que p? Peut-être le fait que la conscience que p ne joue pas exactement le même genre de rôle fonctionnel que la croyance que p. Quoi qu'il en soit, mon incapacité à caractériser le rôle fonctionnel distinctif de la conscience intentionnelle ne saurait être retenu contre moi, puisqu'on ne connaît aucune description du rôle fonctionnel distinctif de quelque type d'état mental que ce soit.

Tournons-nous maintenant vers la conscience intransitive. Selon la conjecture énoncée plus haut, un phénomène mental conscient n'est rien d'autre qu'un phénomène mental dont le sujet qui en est le siège a conscience, au sens de la conscience intentionnelle. Mais cela peut s'entendre de plusieurs manières. L'hypothèse la plus naturelle (mais aussi la plus forte) est sans doute celle selon laquelle si S croit ou désire consciemment que p (c'est-à-dire, s'il croit ou désire que p et si sa croyance ou son désir sont conscients), alors S a conscience qu'il croit ou désire que p (ou de croire ou désirer que p). L'inconvénient de cette hypothèse est qu'elle implique apparemment que nul ne peut avoir de croyances ou de désirs conscients à moins de posséder le concept de croyance ou de désir, ce qui pourrait avoir pour conséquence (i) que seuls les animaux qui (comme certains primates, dit-on) disposent d'une "théorie de l'esprit" peuvent avoir des croyances ou des désirs conscients et (ii) que les bambins ne peuvent avoir de croyances ou de désirs conscients avant d'avoir maîtrisé une "théorie de l'esprit" (soit entre trois et cinq ans, selon les études récentes). On pourrait rétorquer qu'il est de toute manière assez douteux que les animaux (sinon les

bambins) aient des croyances ou des désirs, même inconscients, du moins si avoir une croyance ou un désir consiste à être dans un état interne qui réalise un certain type de rôle fonctionnel. Mais on peut concéder cela sans refuser aux animaux ou aux bambins la capacité d'être dans des états intentionnels (ou à tout le moins représentationnels); le sens de la remarque qui précède est que même dans ce cas on ne pourrait admettre que certains de ces états sont conscients qu'à condition d'admettre aussi que les animaux ou les bambins qui sont dans ces états ont conscience d'être dans des états de ce type, et donc qu'ils ont les concepts de ces types d'états (ce qui exigerait qu'ils aient au moins la capacité d'être dans des états intentionnels d'ordre supérieur).

Quoi qu'il en soit, rien ne nous oblige à nous en tenir à l'hypothèse forte qu'on vient de mentionner. Une autre hypothèse pourrait être que si S croit ou désire consciemment que p, alors il a conscience de sa croyance ou de son désir, ce qui n'exige ni que S ait conscience que l'état interne dont il a conscience est une croyance ou un désir, ni que S ait conscience que cet état interne a pour contenu que p. Entre ces deux hypothèses, il y a place pour une multitude d'autres hypothèses de la forme: si S croit ou désire consciemment que p alors il a conscience qu'il est dans un type d'état qui possède telles ou telles caractéristiques (c'est-à-dire, il a conscience de son état mental et que celui-ci possède certaines caractéristiques). Tous les cas de figures sont possibles en principe, et peuvent, selon les circonstances, être suffisants pour justifier l'affirmation que le phénomène considéré est conscient (au sens de la conscience intransitive).

Il découle de ces remarques que si S a conscience (par exemple) qu'il désire que p (c'est-à-dire, s'il a conscience de désirer que p) alors il désire consciemment que p, mais que S peut désirer consciemment que p sans avoir conscience qu'il désire que p. Tout ce qu'on peut conclure du fait que S désire consciemment que p, c'est qu'il a conscience de son désir. Même en admettant que la conscience intentionnelle de re présuppose la conscience intentionnelle de dicto, il ne s'ensuit pas que S doit avoir conscience que son désir est tel ou tel, mais seulement qu'il doit avoir la capacité d'avoir conscience que son désir est tel ou tel. Et même dans le cas où S aurait en fait conscience que son désir est F (pour un certain F), il ne s'ensuivrait pas que cet F est une caractéristique mentale ou intentionnelle, ni par conséquent que la pensée que x est F est une pensée "d'ordre supérieur"; à moins que toute pensée portant sur un état intentionnel ne soit considérée comme une pensée d'ordre supérieur. Mais ce n'est généralement pas ainsi qu'on entend cette expression; une pensée n'est une pensée d'ordre supérieur, dans le sens visé, que lorsque son contenu ne peut être normalement exprimé qu'au moyen d'un énoncé contenant un contexte indirect (intensionnel). La conception de la conscience intransitive qui se dégage de ces remarques ne

se ramène donc pas à l'idée assez répandue selon laquelle les états mentaux conscients le sont en vertu du fait qu'ils ont une certaine relation avec des pensées d'ordre supérieur.

Il est bon de noter que je ne suis pas en train de proposer que S désire consciemment que p si et seulement si S a conscience de son désir. Le sens de mes observations est qu'avoir conscience de son désir que p est une condition nécessaire mais pas toujours suffisante pour que le désir que p soit conscient, alors qu'avoir conscience que son désir est un désir que p en est une condition suffisante mais non nécessaire. On peut, si on le souhaite, considérer qu'avoir conscience de son désir que p est une condition suffisante pour désirer consciemment que p, mais seulement en concédant que cela ne correspond qu'à une forme minimale de conscience.

Que dire maintenant des qualités subjectives de certains phénomènes mentaux? Selon la conjecture énoncée plus haut, ce sont des qualités des phénomènes mentaux dont le sujet a conscience. Mais si c'est le cas, ce qu'on appelle la conscience phénoménale n'exige strictement rien de plus que la conscience intentionnelle. Car lorsque S a conscience qu'il croit que p, par exemple, il a certainement aussi conscience d'être dans un état qui possède la qualité d'être une croyance que p, et être une croyance que p devrait donc compter comme une qualité subjective de la croyance que p. Bien évidemment, ceux qui parlent de qualia et de feelings bruts et qui s'inquiètent de savoir quel effet ça fait de voir du rouge ou d'avoir mal aux dents n'ont en vue qu'un certain type de qualités subjectives, d'un certain type de phénomènes mentaux, à savoir les qualités "intrinsèques" des vécus ou des expériences. Toutes les expériences sont des phénomènes mentaux conscients. Toutes les qualités d'une expérience dont le sujet a conscience sont des qualités subjectives de cette expérience, mais toutes ne correspondent pas à ce qu'on serait enclin à appeler des qualités intrinsèques de l'expérience, c'est-à-dire des qualia. Ce qui distingue apparemment les qualia des autres qualités de l'expérience dont le sujet a conscience, c'est que ce sont des qualités qu'une expérience ne possède (ou ne peut posséder) que si le sujet en a conscience. En d'autres termes, si un système peut croire que p sans avoir conscience de croire que p, il ne peut apparemment éprouver une douleur sans avoir conscience d'éprouver une douleur, ou tout au moins sans avoir conscience du caractère douloureux de son expérience; et ce pourrait être en cela que consiste le fait que la douleur est une qualité "intrinsèque" de certaines expériences.

Pareille analyse soulève encore une fois, et de manière peut-être plus aiguë, la question de savoir dans quelle mesure il faut avoir le concept de douleur pour pouvoir éprouver de la douleur, ou plus généralement, dans quelle mesure il faut avoir le concept d'une certaine qualité pour pouvoir avoir des expériences qui possèdent cette qualité. Le problème est plus aigu dans le cas des qualités intrinsèques de l'expérience, dans la mesure

où on est plus fermement convaincu du fait que les animaux et les bambins ont des expériences conscientes (qui possèdent des qualités intrinsèques) que du fait qu'ils ont des croyances ou des désirs conscients. Or s'il faut avoir le concept de douleur pour être capable de souffrir, il faut conclure soit que les animaux et les bambins ont le concept de douleur, soit qu'ils ne souffrent pas.

Il semble bien, cependant, qu'on ne possède généralement pas les concepts des qualités intrinsèques de nos propres expériences, et qu'on ne peut, presque toujours, que les décrire de manière indirecte en disant, par exemple, qu'une certaine sensation est une sensation de rouge ou de froid (mais une sensation de rouge n'est pas rouge et une sensation de froid n'est pas froide). Cela suggère que le plus souvent, nous avons conscience de nos expériences et de certaines de leurs qualités, mais sans avoir conscience que ces expériences sont F, où "F" exprimerait le concept d'une qualité intrinsèque (si ce n'était pas le cas, moins de philosophes perdraient leur temps à se demander en quoi consistent les qualités intrinsèques de l'expérience). Certes, on a assez souvent conscience que telle expérience possède une qualité intrinsèque qui ressemble à une des qualités intrinsèques de telle autre expérience; mais il suffit pour cela d'avoir conscience de ces qualités et que ces qualités possèdent certaines caractéristiques, qui n'ont peut-être pas à être elles-mêmes des caractéristiques intrinsèques.

La conclusion que je tire de ces remarques, est qu'un examen plus approfondi de la distinction entre la conscience intentionnelle de re et la conscience intentionnelle de dicto pourrait contribuer à dissiper certaines objections à l'idée que la conscience intransitive et la conscience phénoménale sont des formes de conscience intentionnelle. Cependant, s'il est vrai (comme il y a lieu de le croire) que la conscience intentionnelle de re présuppose la conscience intentionnelle de dicto, en ce sens que nul ne peut avoir conscience de quelque objet que ce soit, à moins d'avoir la capacité d'avoir conscience que cet objet est tel ou tel, alors une condition nécessaire pour que S puisse avoir conscience de ses états mentaux est qu'il puisse avoir conscience que ceux-ci sont tels ou tels (et donc qu'il possède des concepts applicables à ses états mentaux). Or (comme je l'ai déjà remarqué) une condition nécessaire pour que S ait conscience que quelque chose est le cas est qu'il "pense" (dans le sens générique du terme) que cette chose est le cas. Ainsi, même si la conception selon laquelle les états mentaux conscients sont ceux qui causent ou (selon la proposition de Carruthers (1996)) sont aptes à causer certaines pensées d'ordre supérieur doit être rejetée, et même si la capacité d'être dans un état mental conscient n'implique pas la capacité d'avoir des pensées d'ordre supérieur, il ne saurait y avoir d'explication d'aucune forme de conscience qui fasse l'économie d'une explication de la notion de contenu intentionnel. *A fortiori*, il ne

saurait y avoir d'explication naturaliste d'aucune forme de conscience qui fasse l'économie d'une explication naturaliste de la notion de contenu intentionnel.

Il y en a toutefois qui prétendent, non pas que nul ne peut avoir de pensée consciente sans avoir la capacité d'avoir des pensées d'ordre supérieur, mais que

(5) nul ne peut avoir de pensée (tout court) sans avoir la capacité d'avoir des pensées d'ordre supérieur.

Il va sans dire que cette thèse est indépendante de (et donc compatible avec) la thèse selon laquelle

(6) les pensées conscientes sont celles qui causent ou sont aptes à causer des pensées d'ordre supérieur.

D'autre part, en soutenant les deux thèses, on s'engage simplement à soutenir que nul ne peut avoir de pensée à moins d'avoir la capacité d'avoir des pensées conscientes, ce qui semble assez plausible.

En rejetant (6), on ne renonce donc pas nécessairement à (5); or cette thèse fait problème pour le point de vue que j'ai tenté de promouvoir, dans la mesure où elle implique que (contrairement à ce que j'ai suggéré) les animaux et les bambins ne peuvent avoir de pensées conscientes à moins d'avoir la capacité d'avoir des pensées d'ordre supérieur. Il importe donc d'écarter cette possibilité, sous peine de me retrouver devant la nécessité d'avoir à soutenir soit que les animaux et les bambins n'ont pas de pensées du tout, soit qu'ils ont la capacité d'avoir des pensées d'ordre supérieur (et dans ce cas quelle raison y aurait-il de continuer à rejeter (6)?).

Si Davidson ne soutient pas explicitement la thèse (5), l'argument avancé dans "Rational Animals" (1982) pour défendre l'idée que

(7) nul ne peut avoir de croyance à moins d'avoir le concept de croyance

peut donner à penser qu'il serait disposé à la soutenir. Car il est clair, d'une part, que nul ne peut, selon Davidson, avoir de pensée à moins d'avoir des croyances, ce qui veut dire que

(7) est une manière pour lui de soutenir que

(8) nul ne peut avoir de pensée à moins d'avoir le concept de croyance.

D'autre part, la manière dont il défend la thèse (7) est au moins compatible avec l'idée que maîtriser le concept de croyance implique la capacité de l'utiliser en discours indirect (ou l'équivalent "mental"). Il écrit en effet ceci (1982: 479):

Perhaps it is (...) clear that having a belief (...) entails the possibility of surprise. (...) Surprise requires that I be aware of a contrast between what I did believe and what I come to believe. Such awareness, however, is a belief about a

belief: if I am surprised, then among other things I come to believe my original belief was false.¹

Or s'il est logiquement possible d'avoir conscience d'une opposition entre ce qu'on croyait auparavant et ce qu'on croit maintenant sans avoir conscience (par exemple) qu'on croyait auparavant que p et qu'on croit maintenant que q, ou de croire qu'une croyance est fausse sans croire que cette croyance est une croyance que quoi que ce soit, il est au moins naturel de supposer que ce n'est généralement pas le cas. Quoi qu'il en soit, il est clair que si on ne peut avoir la capacité de penser qu'une croyance est vraie ou fausse sans avoir la capacité de penser qu'une croyance est une croyance que p, alors l'argument de Davidson (s'il est correct) ne justifie pas seulement (8), mais aussi (5).

Il est sans doute inutile d'insister sur le fait que "l'argument" en question est loin d'être contraignant, puisqu'il se réduit à l'affirmation qu'il est peut-être clair que si S a des croyances, alors il doit avoir la capacité de croire qu'il a des croyances fausses. Dans ces conditions, on peut aisément rejeter la conclusion soit en observant que l'affirmation en question est vraie, mais ne conduit pas à la conclusion recherchée ("peut-être que p" n'entraîne pas "p"!), soit en niant que S ne peut avoir de croyances à moins d'avoir la capacité de croire qu'il a des croyances fausses.

Il est plus intéressant de remarquer que si le fait d'avoir une valeur de vérité est un trait distinctif des croyances (ou des pensées), ou comme le dit Davidson, si "(m)uch of the point of the concept of belief is that it is the concept of a state of an organism which can be true or false, correct or incorrect" (1982: 479)², alors la position de Davidson pourrait encourager la conclusion qu'aucun état interne d'un système ne peut être vrai ou faux à moins que ce système ne possède le concept de vérité. Autrement dit, pas de valeur de vérité sans concept de vérité. Mais cette interprétation n'est pas la seule possible. Dans un texte plus récent, Davidson (1996: 161) concède qu'on peut dire qu'un tournesol fait une erreur, lorsqu'il se tourne vers une lumière artificielle comme s'il s'agissait du soleil, sans bien sûr aller jusqu'à admettre que les tournesols ont des pensées. Si cela peut être compris comme l'indice que rien ne s'oppose à ce que certains états d'un système aient une valeur de vérité bien que ce système ne maîtrise pas le concept de vérité, alors on peut parfaitement

¹ Dans la traduction française: peut-être est-il (...) clair qu'avoir une croyance (...) implique la possibilité de la surprise. (...) La surprise requiert que je sois conscient du contraste qui existe entre ce que je croyais avant et ce que je viens à croire maintenant. Mais une telle conscience est une croyance au sujet d'une croyance: si je suis surpris alors, entre autres choses je viens à croire que ma croyance initiale était fausse. (1991: 73)

² Dans la traduction française: "Le concept de croyance est principalement le concept d'un état de l'organisme qui peut être vrai ou faux, correct ou incorrect." (1991: 73).

accepter la thèse (5) et nier que les animaux et les bambins aient des pensées ou des croyances, sans pour autant nier qu'ils puissent être dans des états intentionnels vrais ou faux (dont certains pourraient être conscients). Cela revient simplement à admettre que la capacité d'avoir des pensées (ou plus exactement des états intentionnels) d'ordre supérieur trace une ligne de démarcation entre deux types ou niveaux d'intentionnalité (que d'aucuns appellent "proto-intentionnalité" et "intentionnalité", et d'autres "intentionnalité" et "méta-intentionnalité"), et donc de "conscience" intentionnelle. En vue d'indiquer en quoi consiste, selon moi, cette distinction, j'appellerai pour ma part ces deux niveaux celui de la conscience ou de l'intentionnalité individualiste et celui de la conscience ou de l'intentionnalité anti-individualiste. Mais pareille distinction est-elle ultimement viable? Quelles relations y-a-t-il entre l'intentionnalité individualiste et l'intentionnalité anti-individualiste? Et qu'advient-il alors de l'idée que toutes les formes de conscience s'expliquent en terme de conscience intentionnelle?

Je conçois cette distinction comme étant à la fois (i) une distinction entre des systèmes qui ont la capacité d'être dans des états intentionnels de premier ordre sans avoir celle d'être dans des états intentionnels d'ordre supérieur et des systèmes qui ont à la fois ces deux capacités³, et (ii) une distinction "qualitative" entre deux genres de contenu intentionnel (deux manières distinctes d'individuer les contenus intentionnels), de telle sorte qu'il faudrait distinguer, même au niveau des états intentionnels de premier ordre, entre ceux dans lesquels seuls les systèmes anti-individualistes peuvent être et ceux dans lesquels les systèmes individualistes peuvent aussi être (les premiers étant plus finement individués que les seconds).

Un système intentionnel purement individualiste devrait donc aussi être un système qui n'a pas la capacité d'attribuer d'états intentionnels (en mode indirect, ou "sémantiquement opaque") à quoi que ce soit, et qui ne possède aucun concept sémantique ou normatif⁴. On ne doit pas en conclure pour autant qu'un tel système devrait être incapable de réagir différemment aux états mentaux d'autrui, mais seulement qu'il ne pourrait le faire que dans la mesure où ceux-ci se réduisent à des dispositions comportementales. Ainsi pourrait-on dire qu'au lieu d'être totalement dépourvus de "théorie

³ Il doit être clair que nul ne peut avoir la capacité d'être dans un état intentionnel d'ordre supérieur sans avoir la capacité d'être dans un état intentionnel de premier ordre; ou autrement dit, que nul ne peut avoir la capacité de penser que (x pense que p), sans avoir la capacité de penser que q, pour un certain q).

⁴ Je pense aussi qu'un tel système n'est capable ni de délibération authentique, ni par conséquent d'action libre; mais je ne suis pas encore en mesure de justifier cette intuition. Peut-être pourrait-on suggérer qu'on ne peut délibérer sans avoir le concept de raison d'agir et la capacité de distinguer entre de bonnes et de mauvaises raisons?

de l'esprit", les systèmes intentionnels purement individualistes sont tout simplement des behavioristes congénitaux.

2 *Le naturalisme philosophique*

Le naturalisme philosophique est davantage un slogan ou un étendard qu'une doctrine bien articulée. Il procède de la conviction que la totalité des activités humaines, y compris les activités cognitives, sont fondamentalement des phénomènes naturels comme les autres, dont l'étude ne fait pas intervenir d'entités ou de principes radicalement différents de ceux des sciences naturelles. Cette attitude générale est aussi vieille que la philosophie elle-même, et c'est un lieu commun de remarquer qu'elle a connu son heure de gloire au XIXe siècle, sous la forme de ce qu'on a appelé le psychologisme. C'est aussi un lieu commun de remarquer que les deux grands mouvements philosophiques de la première moitié du XXe siècle (la philosophie analytique et la phénoménologie) sont fondés sur la critique du psychologisme et du naturalisme.

On ne peut pas dire que la philosophie de cette période ne contient aucune trace de naturalisme (p. ex. Dewey), mais il ne fait aucun doute que c'est principalement à Quine qu'on doit d'avoir remis la question du naturalisme en haut de l'ordre du jour. Le naturalisme de Quine concerne avant tout l'épistémologie. Dans "L'épistémologie devenue naturelle" (1969a), il soutient qu'il est vain de chercher à proposer des reconstructions rationnelles du savoir, c'est-à-dire de chercher à caractériser des liens logiques entre les données empiriques ou les stimulations sensorielles et les théories que nous acceptons sur leur base, et que la théorie de la connaissance ne peut être autre chose qu'un chapitre de la psychologie empirique (behavioriste).

La phrase précédente pointe en direction des deux aspects du naturalisme que je voudrais distinguer. En disant que l'épistémologie est une branche de la science empirique, Quine ne fait d'une certaine manière que rappeler que l'épistémologie tombe sous le coup de la thèse avancée dans "Les deux dogmes de l'empirisme" (1951), selon laquelle aucun énoncé n'est à l'abri d'une réfutation empirique (c'est-à-dire, aucun énoncé n'est *a priori*). Mais cette suggestion suppose que les énoncés épistémologiques sont des énoncés auxquels on peut, *prima facie*, appliquer la distinction *a priori/a posteriori*. Comme il est raisonnable de penser que seuls les énoncés descriptifs (c'est-à-dire, dans le sens le plus large, ceux qui ont une valeur de vérité) peuvent être dits *a priori* ou *a posteriori*, cela implique que selon cette conception, l'épistémologie est une discipline entièrement descriptive. Et plusieurs commentateurs ont en effet remarqué qu'en adoptant cette position, Quine semble renoncer aux prétentions normatives de l'épistémologie traditionnelle, dont l'ambition est d'identifier les normes qui doivent gouverner la "fixation" des croyances ou des théories, ou en d'autres

termes, d'énoncer des "directives" concernant la formation et la mise à jour des croyances ou des théories.

Mais on aurait tort de conclure, sur la base de ce qu'on vient de dire, que la conception de Quine ne laisse aucune place à quelque espèce de norme que ce soit. Car il reste, théoriquement, la possibilité qu'on soit en mesure de dire (sur le mode assertif/descriptif) quelles sont les normes ("if any") qui, en fait, gouvernent les activités des agents cognitifs (et qu'on ne soit en mesure de le dire qu'*a posteriori*). Et il n'y a, à première vue, aucune raison de penser qu'on puisse le faire pour les normes de la raison théorique sans pouvoir le faire pour les normes de la raison pratique (c'est-à-dire, qu'on puisse avoir une connaissance *a posteriori* des principes de la rationalité épistémique et qu'on ne puisse pas en avoir une des principes de la rationalité prudentielle).

Supposons de plus qu'il y ait de telles normes, alors tout ce qu'on pourrait conclure, sur la base de ce qui précède, c'est qu'il n'appartient pas à l'épistémologue, ou au philosophe, de dire (sur le mode prescriptif/normatif) si nous devons ou non être guidés par ces normes; en d'autres termes, il ne s'ensuivrait pas qu'il est impossible ou inutile de le dire, ou même de l'inférer (mais seulement que de tels jugements normatifs ne relèvent pas de la philosophie/science empirique).

Supposons que nous ayons identifié ces normes, on aurait donc apparemment le loisir d'en conclure que nous devons les accepter, ou que nous ne le devons pas, et dans le deuxième cas, de soutenir qu'il n'y a aucune norme que nous devons accepter, ou que nous devons en accepter d'autres. Aucune de ces recommandations ne pourrait être un énoncé empirique, mais on ne pourrait pas recommander de ne pas accepter ces normes (et donc ne pas les accepter) sans réfuter soit l'hypothèse que nous avons bien identifié les normes que nous acceptons en fait, soit l'hypothèse que la recommandation en question est bien conforme aux normes que nous avons correctement identifiées comme étant les nôtres (ou sans faire la démonstration que ces normes sont incohérentes⁵). Si cette remarque contient ne serait-ce qu'une parcelle de vérité, alors il semble permis de penser qu'au moins certaines formes de naturalisme ne sont pas complètement neutres sur le plan "déontologique", en ce sens qu'elles permettent d'inférer "ce qui doit être de ce qui est". Mais il s'agirait malgré tout de formes assez modestes de naturalisme, dans la mesure où ce lien entre ce qui est et ce qui doit être serait limité aux "faits normatifs" ("si X est ma norme, alors X doit être ma norme"), de sorte qu'on ne pourrait aucunement en conclure que tout ce qui se produit est en accord avec les normes ("si X est le cas, alors X doit être le cas"). D'autre part, comme il y aurait exactement la même relation entre les énoncés normatifs et les énoncés qui décrivent

⁵ Voir Stein 1996: chap. 7 pour le point de vue contraire.

des normes dans le cas où nous pourrions déterminer *a priori* quelles sont les normes qui gouvernent ("en fait") nos activités cognitives, la seule spécificité de ce genre de naturalisme résiderait dans son caractère *a posteriori*.

Cela suggère du même coup que la question de savoir si l'épistémologie (ou, *mutatis mutandis*, l'éthique) est une discipline normative ou non (c'est-à-dire, si elle contient/produit des directives ou des prescriptions) est relativement secondaire, dans le cas où il y aurait des faits normatifs (quelle que soit la manière dont ceux-ci puissent être connus). Et dans le cas où il n'y en aurait pas, il est difficile de voir comment il pourrait y avoir des directives ou des prescriptions capables de guider le comportement.

Tout ceci ne concerne que le cas où il y aurait une place, dans notre "système du monde", pour des énoncés descriptifs concernant les normes gouvernant tel ou tel type d'activités de la part de tels ou tels individus. Or il pourrait s'avérer qu'il n'y a pas de "normes" dans la nature (et donc, du point de vue naturaliste, qu'il n'y a pas de faits normatifs du tout), ce qui voudrait dire que la distinction entre ce qui est et ce qui doit être est vide, c'est-à-dire, que ce qui est coïncide avec ce qui doit être (et non pas seulement que ce qui est la norme coïncide avec ce qui doit être la norme).

En résumé, il semble que le trait caractéristique du naturalisme philosophique consiste à soutenir qu'il n'y a pas de normes non naturelles (ou en d'autres termes, qu'aucune norme n'est connue *a priori*). Lorsque cette thèse minimale s'accompagne de l'idée qu'il y a des normes naturelles, on peut parler de naturalisme normatif (ou modéré, ou réductionniste), et lorsqu'elle s'accompagne au contraire de l'idée qu'il n'y a pas non plus de normes naturelles, on peut parler de naturalisme descriptif (ou radical, ou éliminationniste). Cette terminologie risque toutefois de prêter à confusion, dans la mesure où dans les deux cas c'est du statut d'énoncés descriptifs (à savoir ceux qui décrivent des normes, en affirmant, par ex. que tel ou tel individu accepte ou est guidé par tel ou tel principe) qu'il s'agit, et non pas de la relation entre énoncés normatifs et énoncés descriptifs. Ceci dit cette distinction recouvre assez bien la distinction de Kitcher (1992) entre ce qu'il appelle le naturalisme traditionnel et le naturalisme radical (bien que la discussion de Kitcher soit limitée au naturalisme épistémologique).

Je ne sais pas si le naturalisme est viable, mais je suis relativement convaincu que s'il l'est, cela ne peut être sous la forme du naturalisme normatif qu'à condition d'admettre une ontologie non physicaliste (c'est-à-dire, qu'il y a des faits naturels qui ne sont pas physiques), ou en d'autres termes de supposer que s'il y a des faits normatifs naturels, ce ne sont pas des faits physiques. Cette remarque n'est pas aussi triviale qu'elle paraît, puisqu'on peut certainement être naturaliste sans être physicaliste, et physicaliste sans être naturaliste (en soutenant, par exemple, que la réalité physique épuise la réalité empirique,

mais qu'il y a des faits non physiques accessibles à la connaissance *a priori*). Mais étant donné le physicalisme, il suffit de montrer qu'il n'y a pas de faits normatifs dans le monde physique pour montrer qu'il n'y a pas de faits normatifs.

Il pourrait sans doute, *prima facie*, y avoir toutes sortes de normes; mais il semble plausible de suggérer que s'il y a des normes naturelles, elles sont en premier lieu des normes de rationalité, ou fondées sur des normes de rationalité. Ce sont des principes qui définissent à quelles conditions une action, un désir, une croyance, une intention, ou un raisonnement sont rationnels ou justifiés, et qui concernent donc des états et des processus psychologiques et intentionnels. Or on voit mal comment ces phénomènes intentionnels pourraient être gouvernés par des normes naturelles, s'ils ne sont pas eux-mêmes des phénomènes naturels. D'un autre côté, on ne semble pas avoir autant de mal à imaginer (ou à se convaincre qu'on peut vraiment imaginer) que des normes non naturelles puissent gouverner des phénomènes naturels, ce qui explique pourquoi on a le sentiment qu'il ne suffirait probablement pas de naturaliser l'intentionnalité pour conclure à la vérité du naturalisme philosophique.

Au XIXe siècle, il paraissait évident, et il était largement admis même par les plus féroces détracteurs du naturalisme, que la psychologie intentionnelle était une discipline empirique. La question était de savoir si les principes de la rationalité se ramènent aux lois empiriques de l'enchaînement des états intentionnels, ou s'ils reposent au contraire sur une sorte d'intuition *a priori*. Les naturalistes modérés contemporains posent le problème dans des termes comparables, même si certains (peut-être même la plupart) se donnent la partie plus facile en prétendant non pas que les principes de rationalité sont des lois de la psychologie empirique, mais que les lois et les données de la psychologie empirique sont pertinentes pour déterminer quelles sont les normes de la rationalité (ce qui semble suffisant pour assurer qu'une hypothèse concernant les principes de rationalité de tel ou tel agent est bien une hypothèse empirique)⁶.

⁶ Le moment n'est pas trop mal choisi pour faire un petit commentaire concernant la « querelle » de la rationalité dans les sciences cognitives. Il est bien connu qu'un certain nombre de données expérimentales semblent indiquer que les processus cognitifs humains violent systématiquement les principes de la logique classique et du calcul des probabilités. En supposant que ces principes correspondent bien aux normes de la rationalité (humaine), on peut en conclure que l'expérience montre que les humains ne sont pas rationnels. Ce point de vue semble relever d'une attitude naturaliste, dans la mesure où il revient à dire que la question de savoir si les humains sont rationnels est une question empirique (sans compter que c'est en général un point de vue qui est soutenu par des psychologues); tandis que le point de vue de ceux qui pensent qu'on ne peut pas découvrir empiriquement que les humains sont systématiquement irrationnels semblent au contraire relever d'une sorte d'apriorisme dogmatique. Mais en réalité c'est tout le contraire; cette attitude va de pair avec l'idée que les données empiriques contribuent à déterminer quelles sont les normes de rationalité qui guident un agent, et qu'elles peuvent par conséquent être interprétées comme une réfutation de l'hypothèse que les normes de la rationalité correspondent aux principes de la logique classique et du calcul des probabilités. Évidemment la question demeure, de savoir s'il est possible que les mêmes données empiriques confirment à la fois

Un agent ne peut être gouverné par une norme que s'il a à la fois la capacité de s'y conformer et celle de ne pas s'y conformer. Le premier principe ("ought implies can" ou "à l'impossible nul n'est tenu") est celui sur lequel s'appuient les naturalistes modérés pour soutenir que les normes de la rationalité ne sont pas indépendantes des données empiriques (il est explicitement invoqué par Goldman (1986), *inter alia*). Les deux principes sont essentiels pour éviter la conclusion qu'un agent n'a jamais la capacité de faire autre chose que ce qu'il fait, et le naturalisme radical qui l'accompagne. Mais qu'est-ce qui peut compter comme une raison de croire qu'un agent peut faire quelque chose qu'il ne fait pas, ou ne pas faire quelque chose qu'il fait? Il ne suffit pas, pour montrer qu'un agent a la capacité de faire une chose qu'il ne fait pas, de montrer qu'il y a un monde possible dans lequel il fait cette chose; car il s'agit de savoir s'il aurait pu le faire dans ce monde-ci, à ce moment-là. Et de toute manière, il est difficile de voir comment on pourrait montrer ou réfuter empiriquement qu'une chose qui n'a pas lieu est possible.

Supposons qu'on puisse disposer de cette difficulté, et qu'on ait identifié des principes qu'un certain agent accepte et a la capacité de respecter. Comme on vient de le dire, il doit aussi être possible que cet agent produise un comportement contraire à ces principes. Mais cela peut se produire soit de manière accidentelle (une erreur de jugement ou de raisonnement s'est glissée quelque part; les dispositifs cognitifs n'ont pas fonctionné comme ils auraient dû fonctionner) soit de manière délibérée (l'agent réalise pleinement quelle est l'option rationnelle pour lui, et en choisit une autre). Il est plausible de penser (comme le soutiennent un certain nombre d'auteurs) qu'un agent ne peut être délibérément irrationnel sans l'être aussi accidentellement, en ce sens que le choix d'une option irrationnelle ne peut, ultimement, être rationnellement justifié (mais seulement expliqué causalement). Si cela est juste, il s'ensuit que le principal défi, pour le naturaliste modéré, est celui de faire une place à l'erreur (à l'irrationalité accidentelle), et non à l'incontinence; bien qu'il soit suffisant de faire une place à l'incontinence pour en faire une à l'erreur (en un mot: qui admet l'akrasie ne peut pas être un naturaliste radical).

Peut-on penser qu'inversement, il suffit de faire une place à l'erreur pour en faire une à l'incontinence? Certes, s'il n'y a pas de choix délibérément irrationnel alors tout choix est soit rationnel soit le résultat d'un "accident"; et on pourrait être tenté de suggérer qu'un choix véritable ne peut pas être le résultat d'un "accident". Il ne serait pas raisonnable de céder à cette tentation, car cela risquerait de nous engager dans une régression en nous

l'hypothèse qu'un certain agent accepte certaines normes et l'hypothèse qu'il les viole systématiquement. C'est difficile à dire, mais il semble bien que plus on admet de violations (et plus ces violations sont systématiques, et donc prédictibles), moins il reste de données capables de confirmer l'hypothèse qu'on a identifié les bonnes normes.

obligeant à supposer que tout choix doit lui-même être délibérément choisi, et à admettre que nul ne peut s'écarter des normes sans réaliser qu'il s'en écarte. Mais on pourrait par contre envisager de soutenir que nul ne peut avoir la capacité de s'écarter des normes à moins d'avoir la capacité de réaliser qu'il ne fait pas ce qu'il doit faire (ce qui est une condition nécessaire pour être délibérément irrationnel); car ce n'est qu'une manière de dire que nul ne peut être gouverné par des principes à moins d'être capable de reconnaître les cas qui sont conformes et ceux qui ne le sont pas. Il s'ensuivrait que si un agent a la capacité d'être accidentellement irrationnel, alors il a la capacité de réaliser que certaines options sont irrationnelles. Mais si une option est réellement une option, alors l'agent a la capacité de la choisir, qu'elle soit rationnelle ou non. Il semble par conséquent difficile de supposer qu'un agent puisse avoir la capacité d'être accidentellement irrationnel sans avoir la capacité de l'être délibérément. Il pourrait, autrement dit, s'avérer que si un agent n'a pas la capacité d'être délibérément irrationnel, alors il n'a pas non plus celle d'être accidentellement irrationnel. Si c'était le cas, alors il suffirait de montrer que nul ne peut être délibérément irrationnel pour montrer que nul ne peut être irrationnel, et ainsi établir la vérité du naturalisme radical. L'ironie vient de ce qu'il faudrait alors inclure le Platon du *Protagoras* au nombre des naturalistes radicaux.

Quoi qu'il en soit, les remarques qui précèdent concernant le naturalisme modéré ou traditionnel sont basées sur la prémisse que les phénomènes intentionnels sont des phénomènes naturels, qui est une condition nécessaire pour ce type de naturalisme. Or c'est encore un lieu commun de remarquer qu'une grande partie des efforts récents des philosophes de l'esprit ont précisément visé, sans succès évident, à naturaliser l'intentionnalité, et par conséquent que cette prémisse est loin de pouvoir être prise pour acquise.

Mais qu'est-ce qui permet de douter que les phénomènes intentionnels soient des phénomènes naturels? Une des raisons le plus souvent invoquées (Davidson, Kripkenstein), et celle à laquelle on accorde le plus de poids, semble être que les phénomènes intentionnels (s'il y en a réellement) sont "irréductiblement" normatifs! Mais peut-on réellement offrir cette raison sans commettre une pétition de principe contre le naturalisme modéré, c'est-à-dire sans présupposer qu'il n'y a pas de normes naturelles? En supposant que les phénomènes intentionnels soient normatifs, la seule conclusion qui s'impose, c'est qu'on ne peut pas naturaliser l'intentionnalité sans montrer qu'il y a des normes naturelles, et qu'il serait circulaire de faire usage des concepts de la psychologie intentionnelle (ou de faire référence aux phénomènes intentionnels) en argumentant en faveur du naturalisme (ce qui crée naturellement une difficulté supplémentaire).

Comme il serait vain d'essayer de montrer qu'il n'y a que des normes naturelles en cherchant à montrer qu'il y a des normes naturelles qui gouvernent des phénomènes normatifs, il serait vain (dans cette hypothèse) d'essayer de le montrer en cherchant à montrer qu'il y a des normes naturelles qui gouvernent des phénomènes intentionnels. Cela signifie que s'il y a des normes naturelles, il doit y en avoir qui s'appliquent à un domaine composé de faits ou de phénomènes qui sont non seulement naturels, mais non normatifs. Le naturalisme modéré exigerait alors qu'il y ait des normes qui s'appliquent à des phénomènes non intentionnels (par exemple, qu'il soit "permis" à certains événements physiques d'entretenir telles ou telles relations).

Il semble, en résumé, que le naturalisme en matière d'intentionnalité (c'est-à-dire le naturalisme sémantique) soit une condition nécessaire (mais non suffisante) du naturalisme en matière d'éthique ou d'épistémologie, dans la mesure où les normes éthiques ou épistémiques ne peuvent apparemment s'appliquer qu'à des actions ou attitudes intentionnelles. C'est pourquoi je me tourne maintenant vers la question de savoir si, et à quelles conditions, il est possible d'espérer une explication naturaliste des deux types d'intentionnalité distingués dans la section précédente.

3 Naturalisme et intentionnalité

Un naturaliste sera naturellement enclin à rejeter toute conception interprétationniste (ou projectionniste) des états intentionnels individualistes, c'est-à-dire à penser que les états intentionnels individualistes (ceux des systèmes intentionnels purement individualistes) ont un contenu indépendamment du fait qu'il y ait ou non des systèmes intentionnels anti-individualistes qui ont la capacité de leur en attribuer un. Sa stratégie de prédilection sera donc de chercher d'abord une explication naturaliste de l'intentionnalité individualiste, pour construire ensuite, sur cette base, une explication de l'intentionnalité anti-individualiste. En d'autres termes, il optera pour une analyse "bottom-up", et présupposera que l'intentionnalité anti-individualiste dépend de l'intentionnalité individualiste, mais pas l'inverse.

En admettant, de plus, que la notion de contenu intentionnel est en partie normative, le naturaliste modéré s'engage alors à montrer qu'il y a, dans la nature, des propriétés normatives qui rendent compte des contenus intentionnels individualistes, et d'autres qui rendent compte des contenus intentionnels anti-individualistes. Je pense personnellement qu'une telle attitude est vouée à l'échec, et qu'il faudra ultimement admettre soit qu'il n'y a pas d'intentionnalité du tout dans la nature, soit que la notion de contenu intentionnel n'est pas normative. Les théories naturalistes les plus plausibles, comme celles de Dretske ou de Millikan, qui analysent les contenus intentionnels en termes téléonomiques ou sélectionnistes, ne dépassent guère le niveau de l'intentionnalité individualiste et reposent

d'ailleurs sur une analyse étiologique du concept de fonction qui en évacue la dimension normative. Lorsque les contenus intentionnels individualistes sont analysés de cette manière, ils ne peuvent naturellement pas être aussi finement individués que les contenus intentionnels anti-individualistes (même de premier ordre), puisque (comme les critiques l'ont abondamment souligné) la sélection naturelle ne peut distinguer entre des propriétés nomologiquement équivalentes. Mais le fait que les contenus individualistes soient plus indéterminés que les contenus anti-individualistes (même de premier ordre) ne me semble pas être une raison suffisante pour refuser de les considérer comme des contenus intentionnels, compte tenu du fait qu'ils manifestent (*pace* Heil 1992: 201-202) un certain degré d'opacité sémantique.

Il reste que si on tient à maintenir que la notion de contenu intentionnel individualiste est en partie normative, alors le naturalisme modéré se heurte à la question de savoir si les normes en question peuvent être des normes de rationalité. Dans la mesure où ces normes sont censées être constitutives des contenus intentionnels individualistes, elles ne peuvent être assimilées à des normes de rationalité qu'à condition d'admettre qu'une norme de rationalité peut s'appliquer à autre chose que des états intentionnels déjà constitués. Cette concession ne devrait cependant pas soulever de difficulté, puisque si on est disposé à soutenir avec Davidson que les normes de rationalité sont constitutives des contenus anti-individualistes, cela signifie qu'on est disposé à admettre qu'elles ne s'appliquent pas à des états intentionnels anti-individualistes déjà constitués. Pourquoi dans ce cas insisterait-on pour qu'elles s'appliquent à des états intentionnels individualistes qui seraient constitués indépendamment d'elles? Et pourquoi ne pourrait-il y avoir des normes constitutives de l'intentionnalité individualiste, c'est-à-dire des normes qui s'appliqueraient à des faits naturels non-intentionnels?

Si on admet que les normes constitutives des contenus intentionnels individualistes sont des normes de rationalité, alors on ne peut distinguer les contenus individualistes de premier ordre des contenus anti-individualistes de premier ordre qu'en supposant que seule une partie des normes de rationalité constitutives de ces derniers sont aussi constitutives des premiers (ce qui est au moins une possibilité abstraite). Pareille attitude implique qu'un système intentionnel peut être gouverné par des normes sans avoir la capacité de concevoir ces normes et de les reconnaître comme les siennes, or il semble intuitivement que ce qui fait qu'une norme est une norme de rationalité est précisément (en partie) le fait que les systèmes auxquels elle s'applique ont la capacité de la reconnaître comme telle. Mais on pourrait répondre que cette intuition ne concerne pas le contenu (ou la substance) des normes en question, mais leur statut, et qu'il n'est pas incohérent d'admettre qu'une même norme peut à la fois s'appliquer à des systèmes intentionnels individualistes et n'acquérir le

statut de norme de rationalité que lorsqu'elle s'applique à des systèmes qui ont la capacité de la reconnaître comme une norme (c'est-à-dire à des systèmes anti-individualistes).

Ce qui est relativement clair, c'est que si les contenus intentionnels (qu'ils soient individualistes ou anti-individualistes) sont en partie constitués par des normes de rationalité, il ne peut y avoir d'explication naturaliste de la conscience intentionnelle sans explication naturaliste des normes de rationalité: nature sans raison n'est que ruine de la conscience. Cette conclusion vaut *a fortiori*, dans le cas (quelque peu inorthodoxe) où on préférerait adopter une conception interprétationniste des états intentionnels individualistes, de manière à en faire dépendre l'explication d'une explication (naturaliste) de l'intentionnalité anti-individualiste (plutôt que le contraire). Car dans l'hypothèse où les normes de rationalité seraient constitutives des contenus anti-individualistes, le naturaliste serait toujours engagé à montrer que ces normes existent dans la nature, en plus d'avoir à montrer que ce qui leur confère le statut de normes de rationalité est aussi dans la nature. Mais s'il est vrai qu'une norme n'est une norme de rationalité que si elle est ou peut être reconnue comme telle, alors une explication naturaliste de ce qui fait qu'une norme est une norme de rationalité ne se distinguerait guère d'une explication naturaliste de la conscience intentionnelle (anti-individualiste): nature sans conscience n'est que ruine de la raison. Dans ces conditions, la naturalisation de la conscience et celle de la raison ne seraient que deux facettes d'un seul et même problème.

Puisqu'un système intentionnel anti-individualiste est un système qui possède non seulement des états intentionnels individualistes, mais aussi la capacité de penser que certains de ses états sont des états intentionnels⁷, un tel système doit avoir le concept d'état intentionnel, et il ne peut l'avoir sans avoir le concept de norme (ou de contenu intentionnel) qui va avec. Selon la caractérisation proposée plus haut, il suffit pour être un système intentionnel anti-individualiste d'être capable d'attribuer (en mode indirect) des états intentionnels individualistes (et donc de premier ordre), ce qui exige seulement de posséder le concept de "norme naturelle" et non celui de norme de rationalité. Il semblerait, en d'autres termes, qu'un système anti-individualiste puisse avoir le concept d'état intentionnel individualiste sans pour autant avoir celui d'état intentionnel anti-individualiste. On peut alors se demander s'il ne faudrait pas distinguer deux types de systèmes intentionnels anti-individualistes: ceux qui ont le concept de norme et la capacité d'attribuer des états intentionnels individualistes, et ceux qui ont le concept de norme de rationalité et la capacité d'attribuer des états intentionnels d'ordre supérieur (ou plus généralement, anti-individualistes), ce qui soulèverait la question de savoir ce qu'il faut ajouter à un système

⁷ Ou plus exactement, de penser que certains de ses états sont des pensées que p, pour quelque p.

anti-individualiste du premier type pour en faire un système anti-individualiste du second type.

Mais pareille distinction n'est ni nécessaire ni possible, parce qu'un système intentionnel ne peut avoir la capacité d'attribuer des états intentionnels (c'est-à-dire d'être dans un état intentionnel d'ordre supérieur) sans avoir la capacité d'attribuer des états intentionnels d'ordre supérieur, ni donc avoir le concept de norme sans avoir le concept de norme de rationalité. Il suffit pour s'en convaincre, de se convaincre que nul ne peut avoir la capacité de penser que (X pense que p) sans avoir la capacité de penser que (X pense que Y pense que p)⁸, ou autrement dit, que la capacité de réitérer les attributions d'états intentionnels fait partie de la maîtrise du concept de contenu intentionnel (et cela en dépit du fait qu'un système intentionnel puisse, *pace* Davidson, avoir la capacité de penser que p sans avoir celle de penser que (X pense que p)).

Si cela est vrai, il s'ensuit que nul ne peut avoir la capacité d'attribuer des états intentionnels sans avoir la capacité d'attribuer la capacité d'attribuer des états intentionnels, ou en d'autres termes que nul ne peut avoir la capacité de reconnaître une norme sans avoir la capacité de reconnaître que cette norme peut être reconnue (et qu'elle est donc une norme de rationalité). Il y a donc du vrai dans l'analyse davidsonienne des conditions de possibilité de la pensée: si S est un interprète (c'est-à-dire, s'il a le concept d'état intentionnel), alors S a le concept d'interprète (c'est-à-dire, il a le concept d'état intentionnel d'ordre supérieur); mais cela n'est pas parce que si S est *interprétable* (c'est-à-dire, s'il a des concepts) alors S est un interprète. L'erreur de Davidson est de penser que "l'ascension intentionnelle" commence au sous-sol, quand elle commence au rez-de-chaussée. Je propose, pour clarifier ces remarques, d'examiner brièvement l'argument de la triangulation par lequel Davidson (1989) entend montrer (*inter alia*) que nul ne peut avoir de concepts à moins d'être capable d'en attribuer (et d'avoir le concept de vérité objective).

L'argument consiste à considérer une situation primitive d'apprentissage par conditionnement. Le chien "apprend" à saliver au son d'une cloche (et non plus seulement à la vue de la nourriture), et l'enfant "apprend" à dire "table" en présence d'une table (et seulement en présence d'une table). Comme le note Davidson (1989: 197), en disant que le chien réagit au son de la cloche plutôt qu'aux vibrations de l'air ambiant ou aux stimulations de ses terminaisons nerveuses, nous nous appuyons implicitement sur le fait que les différentes salivations du chien nous semblent suffisamment semblables entre elles pour compter comme des occurrences de la même réponse, et que les différentes sonneries de la cloche nous semblent suffisamment semblables entre elles pour compter comme des

⁸ Voir Heil (1992: 213-214) pour une remarque qui semble aller dans le même sens.

occurrences du même stimulus, tandis que nous ne saurions dire si les processus neurophysiologiques sous-jacents aux différentes salivations du chien sont semblables entre elles ou si les différentes stimulations des terminaisons nerveuses du chien (qui sont causées par les différentes sonneries de la cloche) sont semblables entre elles. Ainsi, pour pouvoir dire à quoi le chien réagit, et comment il réagit, il faut tenir compte de notre propre manière de réagir à la fois aux stimuli qui déclenchent les réponses du chien et à ces réponses elles-mêmes. Les stimuli qui déclenchent les réponses du chien sont des objets que nous trouvons naturellement semblables entre eux (c'est-à-dire, auxquels nous réagissons "de la même manière") et qui provoquent de la part du chien des réponses que nous trouvons naturellement semblables entre elles. Les stimuli sont en quelque sorte les objets ou événements qui se trouvent au point d'intersection de la ligne de perspective du chien et de notre propre ligne de perspective. Selon Davidson (1989: 198), ce genre de triangulation est une condition nécessaire pour qu'il y ait une réponse à la question de savoir à quoi le chien réagit, et par conséquent pour qu'il y ait une réponse à la question de savoir de quoi les concepts d'une créature sont les concepts (dans le cas où elle aurait des concepts):

The problem is not (...) one of verifying what objects a creature is responding to; the problem is that without a second creature interacting with the first, there can be no answer to the question. And if there can be no answer to the question what a creature means, wants, believes or intends, it makes no sense to hold that the creature has thoughts.⁹

Il s'ensuit qu'aucune créature ne peut avoir de pensées à moins qu'une autre créature n'ait la capacité d'interagir avec elle, c'est-à-dire, de réagir à la manière dont elle réagit à son environnement. Davidson entend vraisemblablement suggérer par là qu'aucune créature ne peut avoir de pensées à moins qu'une autre créature n'ait la capacité de lui en attribuer, ce qui est au moins compatible avec une conception interprétationniste des états intentionnels individualistes (certains systèmes pourraient avoir des états intentionnels sans avoir la capacité d'en attribuer, mais ils ne pourraient pas en avoir si aucun système n'avait cette capacité).

Davidson ne s'arrête naturellement pas là. Il ajoute (1989: 199) que pour que les réponses d'une créature à un objet comptent comme des pensées, celle-ci doit avoir le concept de cet objet, et qu'avoir le concept d'un objet, c'est être capable de le situer dans un espace partagé par une autre créature, c'est-à-dire de reconnaître qu'une autre créature est

⁹

Dans ma traduction:

Le problème n'est pas (...) celui de vérifier à quels objets une créature réagit; le problème est que sans une deuxième créature pour interagir avec la première, il ne peut y avoir de réponse à cette question. Et s'il ne peut y avoir de réponse à la question de savoir ce qu'une créature signifie, désire, croit ou vise, il n'y a pas de sens à soutenir que cette créature a des pensées.

capable de réagir au même objet (c'est en quelque sorte posséder le concept de vérité objective). Mais je ne peux voir là qu'une autre manière d'affirmer (sans argument) que nul ne peut avoir de pensées sans être capable d'en attribuer. Ce qui importe, me semble-t-il, c'est qu'en acceptant l'analyse du paragraphe précédent, on a déjà concédé l'essentiel de ce que Davidson semble vouloir soutenir, à savoir qu'il ne pourrait pas y avoir de systèmes intentionnels individualistes s'il n'y avait pas de systèmes intentionnels anti-individualistes. L'intentionnalité individualiste serait donc dérivée et relative, et en attribuant des états intentionnels à des systèmes auxquels nous refusons la capacité d'en attribuer à leur tour, nous leur appliquons des normes exogènes (c'est-à-dire, des normes qu'ils ne partagent pas avec nous, et qui ne sont pas les leurs). Lorsqu'au contraire, nous attribuons des états intentionnels à des systèmes auxquels nous accordons aussi la capacité d'en attribuer à leur tour, nous n'avons d'autre choix que de chercher à leur appliquer les normes qu'ils s'appliquent à eux-mêmes.

Si ce qui précède est sur la bonne voie, on ne peut se demander ni comment un système qui a la capacité d'interpréter autrui pourrait acquérir celle d'interpréter autrui comme un interprète (car l'une ne va pas sans l'autre, quoiqu'il s'agisse bien de deux capacités distinctes), ni comment un système intentionnel individualiste pourrait acquérir la maîtrise du concept de norme et avec elle la capacité d'interpréter autrui et d'être rationnel (car il n'y aurait pas de systèmes intentionnels, ni *a fortiori* de systèmes intentionnels individualistes, si aucun système intentionnel n'avait le concept de norme, c'est-à-dire, s'il n'y avait pas de normes de rationalité). Si on peut douter que tous les systèmes auxquels nous attribuons des états intentionnels en possèdent réellement, il est difficile de douter que certains systèmes ont réellement la capacité d'attribuer des états intentionnels ou qu'il nous semble que certains systèmes ont cette capacité, et *par conséquent* de douter que certains systèmes ont réellement des états intentionnels.

Il n'y aura pas de naturalisation de l'esprit sans naturalisation de cette faculté d'interprétation. Et seul un naturalisme radical peut espérer y parvenir, en soutenant qu'il n'y a pas de faits normatifs, en ce sens qu'il n'est pas nécessaire de postuler l'existence de tels faits pour expliquer quoi que ce soit. Il n'y a de ce point de vue, que deux options possibles: l'éliminationnisme, qui consiste à nier non seulement qu'il y ait, mais aussi *qu'il nous semble y avoir* des phénomènes intentionnels, et le révisionnisme, qui consiste à nier que les phénomènes intentionnels soient essentiellement normatifs.

On tient généralement pour acquis que c'est le naturalisme radical qui a le fardeau de la preuve; et son fardeau n'est certes pas négligeable. Mais il vaut la peine de remarquer que son adversaire n'a guère de quoi se réjouir, puisqu'il lui faut montrer qu'il y a des faits (non normatifs) qui ne peuvent être expliqués qu'en supposant qu'il y a des normes. Or si les

phénomènes intentionnels sont normatifs, il est circulaire de prétendre qu'il faut admettre qu'il y a des normes pour expliquer comment il se fait que nous avons des états intentionnels (cela reviendrait à dire que les normes servent à expliquer les normes). S'ils ne sont pas normatifs, il est difficile de voir en quoi on pourrait avoir besoin des normes pour les expliquer. On ne peut pas prétendre que seule l'existence de normes peut expliquer comment il se fait que nous croyons qu'il y a des normes, ou que nous avons le concept de norme, à moins de supposer que nous ne pouvons pas avoir de croyances fausses ou de concepts vides. Naturellement, on pourrait faire valoir que les normes ne servent pas à expliquer, mais à justifier, et retourner ainsi à la case départ.

La seule chose, peut-être, qu'on peut conclure de tout ce que je viens de dire, c'est qu'on serait mal avisé de se lier les mains par des scrupules naturalistes, ce qui ne veut pas dire qu'on serait mieux avisé de verser dans l'apriorisme, du moins pas sans avoir clarifié davantage la notion d'*a priori*.¹⁰

¹⁰ Une partie de ce chapitre a été présentée le 15 mai 1997 à l'Université du Québec à Trois-Rivières, dans le cadre du congrès annuel de la SPQ, et une autre le 13 mai 1998 à l'université Laval, dans le cadre du colloque *Perspectives sur le fossé explicatif* organisé par Denis Fisette et Pierre Poirier. Je dois des remerciements à Renée Bilodeau, Josée Brunet, Christine Tappolet et Richard Vallée.

