NORMES LOGIQUES ET EVOLUTION

Pascal Engel

Université de Caen
et CREA, Ecole Polytechnique
Revue Internationale de Philosophie 2. 1997,201-219

"The phenomena of reasoning are, in their general features, parallel to those of moral conduct. For reasoning is essentially thought that is under self-control, just as moral conduct is conduct under self-control. Indeed reasoning is *a species of controlled conduct, and as such necessarily partakes of the essential features of controlled conduct. If you attend to the phenomena of reasoning, although they are not quite so familiar to you as those of morals because there are no clergymen whose business is to keep them before your minds, you will necessarily remark, without difficulty, that a person who draws a rational conclusion, not only thinks it to be true, but thinks that similar reasoning would be just in every analogous case. If he fails to think this, the inference is not to be called reasoning. It is merely an idea suggested to his mind and which he cannot resist thinking is true. But not having been subjected to any check or control, it is not deliberately approved and is not to be called reasoning. To call it so would be to ignore a distinction which it ill becomes a rational being to overlook. To be sure, every inference forces itself upon us irresistibly. That is to say, it is irresistible at the instant it first suggests itself. Nevertheless, we all have in our minds certain* norms*, or general patterns of right reasoning, and we can compare the inference with one of those and ask ourselves whether it satisfies that rule. I call it a rule, although the formulation maybe somewhat vague; because it has the essential character of a rule of being a general formula applicable to particular cases. If we judge our norm of right reason to be satisfied, we get a feeling of approval, and the inference now not only appears as irrestisble as it did before, but it will prove far more unshakable by any doubt."*
(Peirce, *Collected Papers*, I, 606)

Je m'accorde pleinement avec Peirce. La logique est une discipline normative. Quand nous disons qu'un raisonnement est correct, ou valide, nous voulons dire qu'il

s'accorde avec certaines règles ou normes du raisonnement correct que nous acceptons. Ces règles ou normes se caractérisent par le fait qu'elles nous paraissent intuitivement évidentes. En ce sens, les règles ou les normes logiques sont comparables, comme le dit Peirce, aux règles de la conduite, aux normes morales. Ce sont des impératifs ou des prescriptions que nous suivons. Mais que sont ces normes ? D'où viennent-elles? Dans la mesure où une norme ou une règle contient un élément impératif ou prescriptif, elle n'est ni vraie ni fausse. On la suit ou ou ne la suit, pas, on s'y conforme ou non. Mais elle ne correspond à aucune réalité. D'un autre côté, les normes et principes logiques semblent aussi être vrais ou faux, correspondre à des faits. Pas des faits empiriques, mais des faits nécessaires. Quand on ouvre un manuel de logique, on ne voit qu'une série de propositions. On nous dit que certaines sont toujours vraies, et que d'autres s'ensuivent en vertu de celles-là. Mais ce faisant on en reste à un niveau purement descriptif, et en ce sens la logique est seulement une description de certaines vérités. D'où la réponse suivante: ce que les normes logiques ont de normatif, elles le doivent à l'existence de faits d'un type spécial, des lois ou les vérités logiques, qui décrivent une réalité particulière, celle des "lois de l'être-vrai", comme disait Frege. La nature fondamentale des règles logiques, selon cette conception, est d'être des vérités indicatives. Mais comment peut-on tirer le "doit" logique d'un "est" logique ? Il y a bien longtemps, et avant Wittgenstein, André Lalande dénonça l'erreur qui consistait à réduire la logique à une science descriptive de faits :

"Un exemple, entre beaucoup d'autres, de cet effort vers le constatif est la célèbre définition de M. Gonseth : "La logique est la science de l'objet quelconque". Formule ingénieuse, et qui ne manque pas de justesse, si l'on n'en fait pas une définition convertible : car elle ne tient compte que des éléments qui, dans les normes de la logique une fois constituée représentent la part de l'objet à connaître, et la "science" qui figure dans le second membre, implique des règles de preuve pour l'établir. Il faut donc la compléter par ce qui donne à celle-ci son autorité : l'obligation de rechercher le vrai et de s'incliner devant une preuve expérimentale dûment administrée." [1]

Mais une fois que l'on admet que la logique est une "science normative" au même titre que les "sciences morales", bien qu'elle ne décrive pas une réalité transcendante de lois nécessaires, comment rendre compte de ces normes ? Que sont-elles si elles ne sont ni des vérités platoniciennes ni des vérités de fait ? La réponse que

[1] A. Lalande, *La raison et les normes*, P.U.F., 2ème édition 1948, p.146-147. Les analyses de Lalande remontent à 1907.

je suis tenté de donner est que les normes logiques font partie des normes de rationalité, que ce sont les règles que doit suivre un agent idéalement rationnel.[2] Mais presque chaque terme de cette définition générale pose problème. Qu'est-ce que la rationalité ? qu'est-ce qu'un agent idéalement rationnel ? La rationalité, au sens, où je l'entends, est-elle même un concept essentiellement normatif. Cela signifie que ce n'est pas un trait qu'un système de croyances ou un agent a en fait, mais un trait qui gouverne notre interprétation d'un système de croyances ou d'un agent, c'est-à-dire la manière dont nous attribuons des croyances et des comportements rationnels à des individus. Si nous ne supposions pas qu'un sujet ou un agent a certains traits de rationalité optimale, nous ne pourrions pas l'interpréter. En ce sens, la rationalité n'est pas une donnée de fait, empirique, que nous pourrions découvrir ou pas chez une créature. Cela ne veut pas dire que nous ne puissions pas, en un sens descriptif, dire que des créatures (humains, animaux, machines) sont plus ou moins rationnelles, plus ou moins irrationnelles. Nous pouvons le dire, mais nos attributions de rationalité sont relatives à des normes de rationalité que nous acceptons. Cela signifie que, quand nous qualifions un système de croyance ou un comportement de "rationnel", nous devons présupposer que ce système ou cet agent *est* rationnel, au sens où il obéit à des normes minimales. Je compte les normes logiques au nombre de celles-ci. Ce sont des normes inférentielles, gouvernant ce que nous pouvons attendre qu'un agent croie, s'il a certaines croyances (en particulier le fait qu'il n'ait pas de croyances contradictoires, et qu'il croie ce qui s'ensuit logiquement de ses croyances). Mais ce sont aussi des normes qui tiennent à la vérité même de ses croyances. Nous devons, en particulier, adopter un principe de charité, d'après lequel l'agent que nous interprétons a des croyances qui sont, d'après nos propres critères, en général vraies. En ce sens, la vérité elle-même n'est pas seulement une propriété descriptive d'un énoncé ou d'une croyance, mais aussi une propriété normative. Le fait même qu'on ait une croyance — qu'on la reconnaisse comme telle — présuppose que l'on *accepte* cette croyance comme vraie, même si cette croyance se trouve, en fait, être fausse. Ce lien intime entre croyance et vérité s'exprime en particulier dans le paradoxe de Moore: on ne peut pas dire "Je crois que *p* mais *p* est faux". C'est en ce sens, notamment, qu'on peut parler d'une norme du vrai.

---

[2] cf. mon ouvrage, *La norme du vrai, philosophie de la logique,* Paris Gallimard 1989, tr.anglaise révisée, *The norm of truth*, Harvester Wheatsheaf, Hemel Hempstead, 1991, en particulier le ch.13.

Selon cette conception, la rationalité d'un système de croyances obéissant notamment à des normes logiques n'est pas un fait; ce n'est ni un fait transcendant au sens fregéen, ni un un fait naturel au sens où l'entend le psychologisme quand il cherche à réduire les "lois de la pensée" à des lois de la psychologie humaine, et ces dernières à des lois naturelles. Dire que les vérités et les règles logiques sont des normes, c'est dire précisément qu'elles ne peuvent pas être déduites, ou réduites à des propositions factuelles portant sur la nature ou la constitution des individus.

Mais est-il vrai que cette conception soit incompatible avec toute forme de psychologisme ou de naturalisme concernant la nature des lois logiques? Il existe une ligne d'argumentation qui permet d'en douter. Admettons, comme le soutient cette conception, que nous ne puissions pas attribuer de croyances à des agents sans supposer qu'ils sont optimalement rationnels. Qu'est-ce qui nous oblige à considérer que cette imputation de rationalité optimale est une condition *a priori* de notre interprétation du comportement, et pas un trait effectif des créatures que nous interprétons? On peut présumer que ce qui nous y oblige est le fait que, précisément, les agents ne sont pas en général rationnels, qu'ils font des erreurs, raisonnent mal, etc., et que par conséquent la supposition qu'ils doivent être rationnels est une supposition de l'interprète, et pas un trait de ce qui est interprété. Mais si l'on pouvait avoir une raison de croire que ce que nous tenons comme des normes de rationalité dans nos interprétations sont *en fait* des propriétés des agents et des systèmes de croyances et d'inférence que nous interprétons, que deviendrait la thèse selon laquelle ces normes de rationalité sont des contraintes conceptuelles *a priori* de notre interprétation ? Elle deviendrait simplement une généralisation factuelle sur nos systèmes de croyance et d'inférence. La raison pour laquelle nous avons telle ou telle norme de rationalité dans l'interprétation refléterait seulement le fait que nos systèmes de croyances et d'inférence *sont*, *as a matter of fact*, rationnels. Or c'est exactement ce que nous serions amenés à dire s'il se révélait qu'*en fait*, la plupart de nos croyances sont vraies (en sorte que le principe de charité ne serait plus un principe, mais l'énoncé d'une loi empirique), et que , *en fait*, nos systèmes de croyances sont non contradictoires et nos systèmes d'inférence aussi parfaits qu'ils peuvent l'être. On répondra que cette supposition est absurde, puisque précisément nous ne sommes pas des êtres idéalement rationnels, logiquement omniscients et capables de dériver de nos

croyances toutes les conclusions qui en dérivent. Mais la supposition n'est pas aussi absurde qu'elle en a l'air, si par "rationalité" d'un système de croyances et d'inférences, nous entendons le fait que ce système fonctionne rationnellement non pas aussi bien qu'il *pourrait* (idéalement) le faire, mais aussi bien qu'il *peut* le faire, et a pu le faire, étant donné les contraintes naturelles qui pèsent sur sa constitution et son évolution. Or telle est bien l'interprétation de la notion de rationalité à laquelle nous conduit l'idée que nos croyances et nos systèmes d'inférence ont été modelés par la sélection naturelle et l'évolution. La sélection naturelle est le mécanisme biologique qui a produit nos systèmes de croyances et d'inférences tels que notre espèce en dispose actuellement. Ces systèmes n'auraient pas été sélectionnés, et par conséquent n'auraient pas survécu, s'ils n'avaient pas pu adapter notre espèce à son environnement de manière à assurer sa survie. Par conséquent, la sélection naturelle doit avoir choisi les meilleurs systèmes de croyance et d'inférence possibles (possibles au sens de l'adaptation maximale aux conditions naturelles de l'espèce, et non pas logiquement ou idéalement possibles). Par conséquent, elle doit nous avoir construits de telle sorte que la majeure partie de nos croyances soient vraies, et le maximum de nos inférences correctes. Un être qui aurait une majorité de croyances fausses, ou qui se tromperait régulièrement dans ses déductions et ses inductions, serait incapable de survivre. Par conséquent, la sélection naturelle nous garantit que la majeure partie de nos croyances sont vraies, et que la plupart de nos stratégies inférentielles sont rationnelles. Nous devons, en ce sens, connaître la logique et atteindre la vérité, ou mourir. Tel est l'argument familier établissant un lien constitutif entre évolution et rationalité.[3] Si cet argument est correct, alors les contraintes normatives de rationalité, et les normes logiques elles-mêmes, ne sont en rien normatives au sens où elles ne peuvent pas dériver de propositions factuelles concernant notre nature: au contraire elles sont factuelles et naturelles au sens où elles sont le produit de l'évolution.

Selon cet argument, les propriétés mêmes de rationalité que nous attribuons aux normes logiques et aux inférences qu'elles justifient sont ce qui permet leur explication naturaliste et évolutionniste. Ne revenons-nous pas alors à ce que Frege et Husserl avaient désigné comme une variété particulière de psychologisme, à savoir

[3] cf. S. Stich, *The Fragmentation of Reason*, MIT Press 1990, ch.3.

l'anthropologisme d'après lequel les lois de la logique sont le produit de l'adaptation optimale de l'espèce humaine, d'après un principe téléologique, et qu'ils avaient critiqué chez des auteurs comme Mach et Avenarius qui formulaient cette adaptation comme une forme du principe de l'économie de pensée?[4] Husserl a très clairement dénoncé ce type d'explication :

> *"La tendance idéale de la pensée logique comme telle la porte à la rationalité. "L'"économie de la pensée"(*sit venia verbo*) en fait une tendance réelle efficace de la pensée humaine, il la fonde au moyen du vague principe de moindre effort, et, en dernière analyse, au moyen de l'adaptation ; et dès lors, il pense avoir ainsi élucidé la norme en vertu de laquelle nous devons penser rationnellement, et il croit, de façon générale, avoir élucidé la valeur et le sens objectif de la science rationnelle. Il est certes, fort légitime de parler de l'économie dans la pensée, d'une "synthèse" économique pour la pensée, de faits au moyen de propositions générales, de généralités inférieures au moyen de généralités supérieures, etc. Mais une semblable manière de parler n'acquiert sa justification que si l'on confronte la pensée de fait avec la norme idéale connue avec évidence, norme qui est ainsi le□□□□□□□□□□□□□□□□□□. La validité idéale de la norme est présupposée chaque fois qu'on utilise avec une signification préalable l' expression d'économie de pensée, par conséquent elle n'est pas un résultat possible d'une explication fournie par la théorie de cette économie. Nous mesurons la pensée empirique d'après la pensée idéale et nous constatons que la première se déroule en fait, dans une certaine mesure, comme si elle était guidée avec évidence par des principes idéaux. Ainsi parlons-nous avec raison d'une téléologie naturelle de notre organisation mentale comme d'un arrangement de celle-ci en vertu duquel nos représentations et nos jugements procèdent grosso modo (c'est-à-dire suffisamment pour la conservation ordinaire de la vie) comme s'ils étaient réglés logiquement. Excepté les rares cas de pensée véritablement évidente, elle ne comporte pas en elle-même la garantie d'une validité logique, elle n'est pas en elle-même et avec évidence ordonnée téléologiquement,ni indirectement, par une évidence intuitive préalable. Mais elle est, en fait, d'une certaine rationalité apparente, elle est telle que nous, théoriciens de l'économie de pensée, quand nous réfléchissons sur les procédés de la pensée empirique, nous devons démontrer à l'évidence que ces méthodes doivent, en général, nous fournir des résultats qui concordent - grossièrement et en moyenne - avec les méthodes rigoureusement logiques... On reconnaît ainsi l'*
> □□□□□□□□□□□□□□□□□□."[5]

Quand Husserl dit que le partisan d'une dérivation téléologique des lois logiques à partir des principes de la théorie de l'évolution et de la sélection naturelle fait un □□□□□□□□□□□□□□□□□□, il veut dire que nous devons déjà connaître l'idéal et la norme avant d'analyser leur fonction dans la connaissance, et, par conséquent, que

---

[4] cf. en particulier Husserl, *Recherches logiques*, I, tr.fr. P.U.F,1959, 2ème ed.1969, p.212 sq.
Mach avait clairement énoncé le principe de cette interprétation téléologique des principes logiques et intellectuels en général : " Une connaissance vraie est toujours un processus psychique qui nous conduit à un résultat biologiquement utile, soit immédiat, soit seulement médiat... Tous les phénomènes de la vie chez l'individu sont des réactions qui s'effectuent pour conserver son existence, et les phénomènes de la vie intellectuelle n'en sont qu'une partie." (*Erkenntnis und Irrtum*, ch. VII, cité par Lalande, op.cit. p.214)
[5] Husserl, *Recherches Logiques* , I, tr.fr. p.230.

nous devons la présupposr pour envisager si la pensée naturelle lui est "conforme". Je crois que Husserl a fondamentalement raison, et que ce qu'il décrit est en fait une version de ce que Rawls, à la suite de Goodman, a appelé la méthode de l'équilibre réfléchi, qu'on peut utiliser pour justifier les lois logiques et développer un agument général en faveur de l'idée que nos systèmes de croyances et d'inférences doivent être présumés rationnels.[6]

On pense souvent qu'il n'est pas nécessaire de réouvrir le dossier du psychologisme et de l'anthropologisme parce que Frege et Husserl auraient démontré une fois pour toutes la fausseté et l'incohérence de ces thèses. Mais ce dossier mérite d'être réouvert, à partir du moment où, comme moi, on définit la logique comme une science normative qui porte sur les croyances d'un être idéalement rationnel. Car les croyances sont des états psychologiques, et nous devons savoir quels liens leurs contenus, même idéaux, entretiennent avec la psychologie des individus. Le dossier doit être réouvert aussi parce que les formes contemporaines du psychologisme et de l'anthropologisme ne sont pas identiques à celles qu'attaquaient Frege et Husserl. J'ai essayé ailleurs de montrer pourquoi une *certaine* forme de psychologisme, appuyée sur les travaux de la psychologie du raisonnement contemporaine, était justifiée, même si l'on adopte une conception normative de la logique.[7] Je voudrais essayer de voir ici en quel sens une certaine forme d'anthropologisme, fondée sur la théorie de l'évolution, pourrait se justifier, et dans quelle mesure.

Il y a trois manières, largement convergentes dans leurs conclusions, mais distinctes par le type d'argument qu'elles utilisent, de chercher à fournir une théorie évolutionniste, donc naturaliste, des normes logiques et de la rationalité de nos systèmes d'inférence.

1) La première est celle indiquée plus haut, c'est-à-dire l'argument qui conduit, à partir du fait que la sélection naturelle choisit les systèmes inférentiels les "meilleurs", à la conclusion que nos systèmes d'inférence doivent être rationnels, et que l'irrationalité est improbable ou impossible. Cet argument repose sur l'idée que la sélection naturelle "optimise" et sur que Gould et Lewontin appellent le "paradigme

---

[6] cf. *La norme du vrai* , op.cit. § 2.3.2, et ch.13.
[7] *La norme du vrai*, ch.13.

panglossien".[8] Il ne permet, cependant, s'il est correct, que d'établir un lien *général* entre la sélection naturelle et la rationalité, et il ne nous dit rien de spécifique sur la nature même et le contenu des normes logiques en particulier.

2) La seconde manière consiste à défendre une théorie générale des contenus intentionnels d'après laquelle une croyance donnée a le contenu qu'elle a en vertu du fait qu'elle remplit une certaine *fonction* biologique. Le caractère normatif des contenus mentaux en général, et des contenus de pensée logiques en particulier, s'explique donc, selon cette conception fonctionnaliste-téléologique, par le fait que ces contenus ont rempli, et remplissent encore, dans l'espèce une certaine fonction. Diverses variétés de "téléosémantique" ont en ce sens été proposées.

3) La troisième manière consiste à essayer de dériver, par une histoire biologique ou évolutionniste, le contenu de certaines capacités inférentielles et de certains principes logiques. On essaie ainsi, en étudiant les propriétés de notre raisonnement naturel sur certains contenus logiques, de montrer en quoi ces propriétés s'expliquent mieux si l'on suppose qu'ils sont le produit de schèmes mentaux hérités de l'évolution de l'espèce.

Chacun des arguments mérite un examen détaillé. Mais je ne pourrai ici que les analyser dans leurs grandes lignes pour en repérer les difficultés.

1. L'argument du lien général entre évolution et rationalité

J'ai déjà énoncé ci-dessus les grandes lignes de l'argument. Il faut le préciser. L'idée de base est que l'évolution produit des organismes dont la construction, le *design* , est optimal. On exprime cela en général en termes d'adaptation ou de "*fitness*". Mais ces notions sont notoirement ambiguës et difficiles.[9] Au sens le plus général, la *fitness* d'un organisme, c'est sa capacité à survivre, au sens spécifique et contemporain, c'est le succès reproductif (mesuré quantitativement) d'un trait. On dira qu'un organisme est mieux construit qu'un autre s'il maximise sa *fitness* plus que cet autre (ce qui pose le problème de savoir comment on peut comparer ces organismes, en particulier s'ils

[8] cf. Dennett, "Intentional Systems in Cognitive Ethology", *Behavioral and Brain Sciences*, 6, 1983, repris dans *The Intentional Stance*, MIT Press, 1987, tr.fr. P. Engel, *La stratégie de l'interprète*, Paris, Gallimard 1990

[9] cf. E. Sober, *The nature of evolution*, MIT Press 1984, et les commentaires de J. Gayon, "Epistémologie du concept de sélection", *L'Age de la Science*, II, 1989, 201-227. Je m'appuie dans tout ce qui suit sur Stich, op.cit., ch.3

vivent dans des environnements complètement différents, et *a fortiori* s'ils appartiennent à des espèces différentes). On passe de là à l'idée qu'un système cognitif bien adapté ou *fitted* est rationnel. Mais si l'on esssaie de préciser l'argument, on s'aperçoit qu'il est très fragile.

Tout d'abord que veut dire, dans ce contexte, "rationnel" ? On peut considérer que cela veut dire seulement "bien adapté". Mais c'est bien vague et général, et cela court le risque de commettre le fameux raisonnement tautologique que l'on reproche habituellement aux explications par la sélection naturelle. Si la sélection naturelle est la capacité des plus aptes à survivre, et si l'on définit les plus aptes comme ceux qui survivent, on voit mal comment on peut apprendre quoi que ce soit de ce type d'explications. De même si on dit qu'un système est rationnel s'il est bien adapté au sens où il est le produit de l'évolution par sélection naturelle, on n'avance rien d'informatif, puisqu'on dit seulement qu'un système qui est le produit de l'évolution est le produit de l'évolution. On peut dire qu'un système cognitif optimise sa *fitness* s'il produit plus de croyances vraies qu'un autre, et évite mieux les erreurs. En retour les systèmes d'inférence d'un organisme sont plus "*fit* "s'ils produisent plus de croyances vraies. Supposons que cela vaille pour les croyances logiques. Par définition ces croyances, si l'on dispose d'un système inférentiel approprié, doivent permettre d'en inférer une infinité d'autres. Et même une contradiction peut entraîner une infinité de propositions, parmi lesquelles une infinité de tautologies. Par conséquent, l'idée qu'un système est "*fitted*" s'il produit le plus de croyances vraies est quasiment triviale si elle s'applique à un système qui possède des règles de logique déductive. La thèse n'a vraiment d'intérêt que si elle porte sur des croyances empiriques particulières et inductives, qui, par définition, n'ont pas le pouvoir d'entraîner automatiquement d'autres croyances vraies. Limitons donc la thèse à ce type de croyances. Il est facile de produire des contre-exemples dans lesquels des croyances fausses et des systèmes d'inférence peu fiables conduisent à des stratégies optimisatrices pour l'adaptation et la reproduction. Par exemple, Stich[10] a évoqué une expérience dans laquelle des rats sont rendus malades par des doses de radiations après avoir reçu une certaine nourriture qui a une certaine odeur. Les rats acquièrent la croyance fausse que la nourriture est empoisonnée, mais se portent indubitablement mieux s'ils acquièrent

[10] Stich, op.cit.p.62.

cette croyance. Un autre exemple que donne Stich est celui de la pléiotropie, dans lequel un seul et même gène affecte deux traits ou systèmes. Il se produira parfois qu'un seul et même gène ait des effets positifs sur un système et des effets négatifs sur l'autre. Dans ces cas la situation optimale serait sans doute de garder le positif et d'éviter le négatif. Mais il y a nombre de cas où le trait positif et le trait négatif persistent. Par exemple, les animaux qui vivent dans l'Arctique, comme les ours blancs, ont les gènes de l'albinisme, qui produisent des pelages blancs. Ceux-ci ont une valeur adaptative évidente dans ces contrées, mais produisent aussi des difficultés visuelles chez les animaux en question, et par conséquent un inconvénient adaptatif évident. Dame Nature aurait mieux fait de produire un albinisme sans problèmes visuels, mais apparemment elle n'en a pas eu l'idée. Stich objecte encore que le fait qu'un trait soit dominant dans une population ne signifie pas pour autant qu'il optimise la *fitness*, et que la distribution de ce trait peut varier de façon tout à fait indépendante de son succès reproductif. Il n'y a pas de raison de considérer, par exemple, que les variations dans les langues que parlent les populations soient liées systématiquement à des avantages reproductifs, ou que la prédominance d'une langue comme l'anglais le soit.

Je ne suis pas sûr, cependant, que ces arguments contre l'argument liant évolution et rationalité soient très convaincants. Par exemple, pour prendre le cas du rat qui survit grâce à des croyances fausses, il n'est pas sûr que ce soit les croyances du rat qui soient ici en cause, mais son comportement. L'évolution a plutôt tendance à sélectionner des comportements que des croyances, même si l'on considère que les croyances sont des composantes des comportements. Mais surtout, et d'une manière générale, le partisan de l'argument évolutionniste dont il est ici question peut toujours répondre que tout dépend de la manière dont on découpe les événements, les périodes temporelles, et par suite de la manière dont on calcule les gains et les désavantages reproductifs. Dans certaines circonstances, il peut être adaptativement avantageux de suivre une procédure qui semble irrationnelle ou d'avoir des croyances fausses: mais rien n'empêche de considérer cette procédure comme globalement irrationnelle au sens très vague où elle est "bonne", alors qu'elle est localement irrationnelle si on la considère de façon isolée. De même l'argument selon lequel des croyances fausses peuvent se révéler utiles n'a aucun poids. La prémisse de l'argument

liant l'évolution à la vérité des croyances n'est pas que *toutes* les croyances d'un organisme qui a évolué sont vraies, mais que *la plupart* doivent l'être, et qu'un organisme qui a été sélectionné par l'évolution a plus de chances d'avoir des croyances vraies que des croyances fausses. Il est frappant de constater que la même objection a été faite au principe de charité interprétative défendu par des auteurs comme Davidson et Quine. Ce principe, nous dit-on, ne peut ni être un principe qui décrit notre pratique usuelle d'interprétation des croyances et des énoncés linguistiques, ni un principe normatif, parce qu'il rend impossible quelque chose qui se produit en fait tout le temps, à savoir que nous nous trompons, et parce qu'il interdit de dire que nous pourrions découvrir des personnes qui seraient irrationnelles. Mais le principe de charité n'interdit rien de ce genre. Il dit seulement qu'il y a peu de chances pour que nous découvrions que des individus se trompent dans *toutes*, ou même la plupart, de leurs croyances, et que nous puissions découvrir que des individus sont *totalement* irrationnels. Pour les attributions les plus usuelles, le principe ne fonctionne pas à vide: il ne fonctionne que relativement à d'autres hypothèses que l'interprète fait sur le comportement, les croyances et l'environnement du sujet[11]. Il n'est pas, en d'autres termes, le seul principe d'interprétation. Il en est de même du principe supposé de sélection des croyances vraies de l'argument qui nous intéresse. Dire que l'évolution sélectionne les systèmes inférentiels qui ont le plus de chances de produire des croyances vraies ne veut pas dire qu'elle sélectionne des organismes infaillibles ou incapables de se tromper. Mais du même coup, cela prive le type d'explication recherché de toute véritable portée parce que ce qui nous intéresse, dans une explication de ce genre, c'est la possibilité d'établir une corrélation systématique entre la production, par un système inférentiel, de vérités, et son succès adaptatif. Mais précisément, comme on va le voir, c'est ce genre de corrélations systématiques qui fait défaut.

2. La conception téléosémantique des contenus mentaux

Tournons-nous alors vers une tentative reposant sur le même genre d'argument général liant l'évolution à la rationalité, mais cherchant néanmoins à reconstruire plus

[11] cf. en particulier les remarques de Daniel Laurier, "Rationality and Intentionality, a Defense of optimization in theories of intentionality", à paraître

précisément le lien entre les *contenus* des croyances et leur fonction sélective. Rappelons que ce que nous cherchons, c'est une théorie qui nous permette de dériver systématiquement les contenus de pensée "logiques" de leurs fonctions biologiques. Ceci veut dire, notamment que nous soyons capables, en particulier d'expliquer le sens de phrases où figurent des constantes logiques telles que "et", "ou", "ne... pas", "si...alors" ou "quelque" en termes du rôle fonctionnel que jouent ces mots au sein du système cognitif d'un individu. Le fait que la thèse porte sur les *contenus* de pensée , et non pas sur les mécanismes qui produisent ces contenus, est fondamental. La plupart de ceux qui s'opposent à une explication évolutionniste des lois logiques ou de la pensée en général ne nient pas que les mécanismes et les processus qui produisent ces lois et ces pensées peuvent être et sont sans doute le produit de l'évolution, au même titre que nos organes. Ce qu'ils rejettent est l'idée que l'on pourrait donner une explication biologique des contenus mêmes des pensées et des représentations. Or c'est précisément cette idée que proposent les théories de la représentation mentale contemporaines que l'on appelle "téléologiques" ou "téléosémantiques".[12] D'après ces théories le contenu d'un état mental, et la signification d'une phrase qui l'exprime, est précisément la fonction que joue ce contenu, au sens biologique du mot "fonction". Le contenu d'un état mental ou d'une phrase est déterminé par ses conditions de vérité, qui sont les circonstances dans lesquelles le contenu a conduit à des actions couronnées de succès pour l'organisme qui entretient ce contenu. Ainsi une croyance (un type de croyance) a un certain contenu si elle a une certaine fonction, et elle a cette fonction si dans le passé un nombre suffisant de ses occurences ou *tokens* ont rempli cette fonction dans le passé. Or une croyance ne remplit une certaine fonction "normale" dans un organisme que si elle est vraie, c'est-à-dire a eu un avantage sélectif par la fonction qui lui correspond. C'est ainsi que l'on retrouve le lien présupposé par l'argument (1) précédent entre l'évolution et la vérité des croyances. Ce qu'il y a d'intéressant dans cette conception est l'idée qu'une croyance a une certaine fonction "normale", définie en termes téléologiques. Une fonction est une propriété du second-ordre d'un certain mécanisme causal (par exemple un mécanisme de reconnaissance visuelle), qui rapporte ce mécanisme à un environnement. Dire que la fonction est

---

[12] En particulier Millikan, *Language, Thought, and other biological categories*, MIT 1984 ; Papineau, *Reality and Representation*, Blackwell, 1987, Mc Ginn, *Mental Content*, Blackwell, Oxford 1989., et F. Dretske, *Explaining Behaviour*, MIT Press , 1988. cf. P. Engel, *Etats d'esprit*, Alinéa, 1992, ch.5

normale, c'est dire que le mécanisme fonctionne correctement. Il est clair en ce sens que la notion de fonction normale est une notion normative; mais c'est une notion normative formulée en termes naturalistes, en termes de la notion de "bon fonctionnement d'un mécanisme naturel". En ce sens un raisonnement "correct" est un raisonnement qui est le produit du bon fonctionnement d'un certain dispositif inférentiel. On peut espérer ainsi réduire les normes dont on dit qu'elles sont propres aux contenus de pensée en général, et aux contenus logiques en particulier, à des normes naturelles, c'est-à-dire à des fonctions. Un autre avantage de la notion de fonction télélologique est de permettre d'expliquer, du moins de prime abord, le problème qui affecte toute théorie naturaliste de la représentation, à savoir celui de l'erreur. Si le contenu d'une pensée est, comme le dit tout théorie de ce type, constitué par les circonstances régulières de l'environnement qui la rendent vraie, comment peut-il y avoir des pensées fausses, des pensées à propos de ce qui n'existe pas ? Si l'on dispose de la notion de fonction normale, le problème est apparemment résolu si l'on dit que l'erreur est un certain type de dysfonctionnement.

Cette thèse se heurte à de nombreuses difficultés, que je ne peux examiner ici[13]. Mais il y en a deux principales. La première est que l'on ne peut spécifier qu'une fonction est "normale" que si nous supposons que les conditions dans lesquelles elle s'exerce pour l'organisme considéré sont elles-mêmes "normales". C'est un peu le même problème que celui du rat évoqué tout à l'heure. Son dispositif de détection de nourriture fonctionne normalement, mais il le conduit à avoir des croyances fausses. Par contre les conditions dans lesquelles le dispositif s'exerce ne sont pas normales, puisque l'on associe artificiellement une bonne nourriture à des radiations malfaisantes pour le rat. Quelles sont alors les conditions normales de fonctionnement du dispositif? La seconde difficulté est qu'il est très difficile de dire quelles sortes de croyances sont associées à des actions réussies en termes d'avantage sélectif, et donc de dire quelles sortes de fonctions sont ainsi sélectionnées. [14] Comment peut-on individualiser suffisamment tel ou tel type de croyance indépendamment d'autres, en

---

[13] cf. le livre mentionné ci-dessus, ch. V, et mes articles "Teleosemantics, Realist or Anti-realist?" in H.J. Sandkühler, Hrg, *Wissenschaft und Wissen*, P. Lang, Berne, 1991, "Intentionalité, interprétation et téléologie", à paraître in D. Janicaud, dir., *L'intentionalité*, Actes du colloque de Nice, Juin 1992, ainsi que D. Laurier, "Pangloss, l'erreur et la divergence", in D. Andler et alii, *Epistémologie et cognition*, Bruxelles, Mardaga, 1992

[14] (cf. le cas de la bactérie de Dretske : "croit"-elle quelque chose au sujet du Pöle magnétique ou au sujet de la direction d'une eau riche en oxygène? cf. des difficultés semblables analysées par Dennett, dans le ch. VIII de *The Intentional Stance.*

particulier dans le cas des systèmes cognitifs humains qui manifestent un fort degré de holisme (quoi que l'on veuille dire par ce terme)? Dans un cas comme dans l'autre, la difficulté à laquelle nous nous heurtons ressemble fortement à celle de l'□□□□□□□□□□□□□□□□□ dénoncé par Husserl : ne faut-il pas déjà avoir spécifié quel est le contenu de telle ou telle représentation mentale pour pouvoir déterminer quelle est sa fonction normale en termes évolutionnistes, au lieu du contraire ? Si la stratégie naturaliste en question ne parvient même pas à individualiser des contenus mentaux élémentaires, comme les contenus perceptifs, comment peut-elle *a fortiori* y parvenir pour des contenus plus complexes comme ceux qui mettent en jeu des notions logiques ?

3. La dérivation évolutionniste des normes logiques

Il y a un troisième type de tentative de "darwiniser" la logique. Elle consiste, comme la précédente, à expliquer le caractère normatif de ses règles en termes naturalistes. Mais à la différence de la précédente, elle ne cherche pas à expliquer les contenus spécifiques de ces règles. Elle se contente d'essayer de montrer que notre capacité même à avoir des normes logiques est le produit de l'évolution, et non pas que les produits particuliers de l'exercice de cette capacité sont eux-mêmes de tels produits. Le pouce du Panda est le produit de l'évolution, mais quand le Panda se gratte l'oreille avec son pouce à tel moment, il n'y a pas d'explication évolutionniste particulière de ce grattement particulier.

Il faut revenir au rapprochement que faisait Peirce entre les normes logiques d'une part, et les normes morales d'autre part. Nous pouvons, au sujet des secondes, poser la question : quelle est leur fonction, et comment en sommes nous venus à les accepter ? La réponse fournie par différents auteurs récemment, et notamment celle de Allan Gibbard dans son livre *Wise choices, Apt Feelings* [15] est la suivante. Accepter une norme —la considérer comme telle— est un état psychologique. Cet état psychologique est une capacité spéciale, qui fait partie du système de motivation et de contrôle des humains, et qui dépend du langage. Cette capacité est le produit de l'adaptation humaine. Sa fonction, selon Gibbard, est une fonction de *coordination*..

[15] Harvard. U. Press, 1990, tr.fr. à paraître, P.U.F.

Tout comme nous, nos ancêtres lointains se trouvèrent dans des situations où la survie et la reproduction étaient favorisées par la coopération avec les autres. Il faut comprendre nos capacités normatives comme facilitant la coopération, ou la coordination, et ainsi comme augmentant la *fitness* de l'espèce.

L'idée sur laquelle s'appuie Gibbard est empruntée à la théorie des jeux. La structure de l'intéraction humaine est celle d'un problème de "marchandage" (*bargaining*). Supposez deux joueurs A, et B, qui ont le choix, dans une situation donnée, entre deux options : coopérer ou ne pas coopérer pour obtenir un certain avantage. La coopération requiert deux ensembles d'attentes coordonnées, des attentes portant sur ce que l'autre fera pour obtenir un résultat favorable aux deux et des attentes portant sur la division des fruits de la coopération. Qu'il faille coopérer est illustré par les cas familiers de "dilemme du prisonnier":

		joueur A	
		C	non C
joueur B	C	4, 4	0,5
	non C	5, 0	2, 2

Ce jeu est paradoxal car quoi que fasse B, il vaut mieux pour A ne pas coopérer, et quoi que fasse A, il vaut mieux pour B ne pas coopérer. Par conséquent, en ce sens, il est rationnel (i.e cela maximise l'utilité subjective) de ne pas coopérer, et pourtant chacun des deux joueurs trouverait avantage à coopérer. La bonne stratégie est ici de ne pas coopérer. Mais supposons que les joueurs répètent ce jeu 10 fois. Il peuvent avoir deux stratégies : coopérer ans le premier jeu, puis faire ce que fait l'adversaire au coup précédent ("donnant donnant"). La matrice devient alors:

	C	PPR
C	20	23
PPR	18	40

Dans ce cas, chacun des deux joueurs, quand ils coopérent, tirent bénéfice de la coopération. Des biologistes comme Maynard Smith[16] ont proposé sur ces bases une "théorie des jeux évolutionniste" appliquant ces principes à la sélection naturelle et aux situations conduisant à des strétagies évolutives stables, c'est-à-dire qui favorisent l'adaptation. On peut alors suggérer que les normes, les conventions morales que nous adoptons et acceptons sont les résultats de jeux de coordination de ce genre. Les contrats sociaux même que nous passons sont des effets de ces coordinations[17]. Selon Gibbard, nos systèmes moraux sont le produit de cette évolution de la coopération entre humains. Gibbard adopte et développe une théorie expressiviste et subjectiviste de l'éthique: nos jugements moraux et nos valeurs morales sont l'expression de nos sentiments et de nos émotions, et des efforts de coopération que nous faisons pour coordonner nos émotions avec celles des autres. Plus généralement, il dit qu'"appeler quelque chose rationnel c'est exprimer notre acceptation d'une norme qui l'autorise." Ainsi les normes sont-elles à la fois subjectives (parce qu'elles sont le produit de sentiments moraux, au sens de Hume ou d'Adam Smith) et objectives, parce qu'elles sont des projections de nos systèmes d'attentes et de coordination. Or, ne peut-on pas étendre ce que dit Gibbard des normes morales aux normes logiques? Certes, il est beaucoup plus difficile de dire que les normes logiques sont le produit de l'évolution de nos sentiments et de nos émotions que cela ne peut l'être pour le cas des normes morales. Mais on peut émettre la conjecture plausible selon laquelle notre *acceptation* de certaines normes logiques est le produit de nos efforts de coopération. Tout ceci suppose déjà une première strate d'évolution, l'évolution du langage. Nous ne pouvons nous coordonner avec autrui que parce que nous pouvons parler, entrer en conversation et communiquer avec les autres. L'hypothèse en question ne cherche donc pas à dériver des contenus de jugements logiques particuliers, par exemple des jugements négatifs, conjonctifs ou disjonctifs, de capacités naturelles. Mais elle soutient que la manière dont nous sommes venus à *accepter* , à trouver rationnels, des principes ou des normes tels que le principe de non contradition ou du tiers exclu est le produit d'un tel effort de coopération. Contrairement au psychologisme ou à l'anthropologiqme du XIXème siècle de Bain

---

[16] J. Maynard-Smith, *Evolution and the Theory of Games* , Cambridge University Press, 1975

[17] cf. également R.Axelrod,*The Evolution of cooperation,* New York, Basic Books, 1986, tr.fr. *Donnant donnant,* O. Jacob, Paris, 1992

ou de Mill, qui essayait de dériver les contenus logiques de certains actes de l'esprit, l'anthropologisme contemporain ne se prononce que sur l'état psychologique d'acceptation de normes, c'est-à-dire d'internalisation des attentes coordonnées. Pour évaluer complètement la thèse de Gibbard, il faudrait évaluer sa théorie expressiviste des normes elle-mêmes, ce que l'on ne peut faire ici. Supposons cependant qu'elle soit correcte. Mais en ce cas, elle est si générale que l'on ne peut dire qu'on a en quoi que ce soit " naturalisé" les normes logiques.

Dans quelle mesure peut-on aller plus loin, c'est-à-dire dans quelle mesure peut-on dire que les contenus spécifiques de jugements logiques particuliers sont également le produit de jeux coopératifs évolutionnistes ? Certains psychologues se sont risqués à cette hypothèse, au sujet du cas particulier des jugements conditionnels de la forme "si ... alors", étudiés à travers la fameuse "tâche de sélection" de Wason[18]. Selon ces auteurs, les performances inférentielles des sujets qui manipulent des raisonnements conditionnels courants sont fortement améliorées si l'on suppose que les règles qu'ils suivent reposent sur des "contrats sociaux" de la forme "Si tel bénéfice, alors tel coût". Ces "contrats sociaux" seraient des "algorithmes" darwiniens ayant évolué, et les sujets raisonneraient d'autant mieux qu'ils suivraient ces algorithmes. Mais tout ce que l'on peut dire, c'est que les capacités inférentielles en général, tout comme les capacités linguistiques, des individus sont des produits de l'évolution, et qu'en ce sens elles sont rationnelles. Mais rien, dans les travaux psychologiques et les hypothèses d'algorithmes darwiniens ne permet de dire que tel système inférentiel, telle norme logique plutôt que telle autre, est le produit de l'évolution. C'est de l'aptitude générale à inférer et à penser logiquement que l'on peut dire qu'elle est le produit de l'évolution, pas des schèmes d'inférence particuliers. Si c'est le cas, alors la thèse d'une "réduction" évolutionniste et naturaliste des normes logiques peut bien être correcte, mais elle est si générale qu'elle n'a aucun effet explicatif.

Revenons à Peirce. Il comparait la normativité des règles logiques à la normativité des règles morales, et celle de ces dernières à la normativité de certaines conduites. Il envisageait que les normes logiques, comme celles de la conduite, soient le produit de certaines habitudes mentales ou dispositions, elles-mêmes produites par l'évolution.

---

[18] cf. les références données au ch. XIII de mon livre cité en note 2, et P.Engel, "Logique, raisonnement et rationalité", à in O.Houdé et D. Miéville, dir., *Philosophie de la pensée logico-mathématique,* Paris, PUF, 1993.

En ce sens, nos croyances logiques, c'est-à-dire nos croyances du second-ordre au sujet de nos croyances — nos "leading principles"— peuvent bien être le produit de l'évolution, et provenir des sentiments d'approbation que nous éprouvons à la la suite d'une longue histoire interactive avec nos semblables (qui les ont approuvées). Mais si cette capacité *générale* à avoir des systèmes référentiels et à les approuver (à les considérer comme des normes) a évolué, il ne s'ensuit pas que le *contenu* des règles inférentielles elles-mêmes soit dérivable d'une histoire évolutionniste. A partir du moment où ces règles énoncent certains idéaux de pensée et d'action, leur origine devient opaque,et c'est cette opacité même qui est le signe de leur normativité et leur rationalité.*

* Des versions antérieures de ce texte ont été lues à l'Université de Rennes en janvier 1992 et à l'Université de Rouen en octobre 1992.