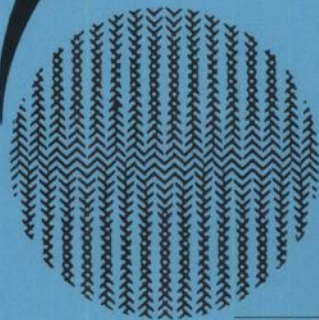


Jean-Pierre Dupuy

**Les savants croient-ils
en leurs théories ?**

**Une lecture philosophique
de l'histoire des sciences
cognitives**



INRA

EDITIONS

Jean-Pierre Dupuy

Les savants croient-ils en leurs théories ?

**Une lecture philosophique
de l'histoire des sciences
cognitives**

Une conférence-débat organisée
par le groupe *Sciences en questions*
Paris, INRA, 15 janvier 2000

Institut National de la Recherche Agronomique
147 rue de l'Université, 75338 Paris cedex 07

La collection "Sciences en questions" accueille des textes traitant de questions d'ordre philosophique, épistémologique, anthropologique, sociologique ou éthique relatives aux sciences et à l'activité scientifique. Elle est ouverte aux chercheurs de l'INRA mais aussi à des auteurs extérieurs.

Directeurs de collection
Etienne Landais, Raphaël Larrère

Le groupe de travail "Sciences en questions" souhaite favoriser la réflexion critique des acteurs de la recherche sur l'activité scientifique et ses implications. Son ambition première est d'enrichir la réflexion interne à l'INRA en l'alimentant de contributions propres à éclairer, sous une forme accessible et attrayante, les débats contemporains sur la science et la recherche.

Texte mis en forme par l'auteur, avec la collaboration d'Etienne Landais, à partir de l'enregistrement de la conférence.

© INRA, Paris, 2000 ISSN : 1269-8490 ISBN : 2-7380-0944-1

Le code de la propriété intellectuelle du 1^{er} juillet 1992 interdit la photocopie à usage collectif sans autorisation des ayants droit. Le non-respect de cette proposition met en danger l'édition, notamment scientifique. Toute reproduction, partielle ou totale, du présent ouvrage est interdite sans autorisation de l'éditeur ou du Centre français d'exploitation du droit de copie (CFC), 20 rue des Grands-Augustins, 75006 Paris, France.

Préface

C'est une épreuve redoutable, que de présenter un ami. Lorsque tu m'as envoyé ta bibliographie, j'ai pris conscience, en outre, que les trois livres et les quelques articles que j'avais lus ne sont que des fragments qui jalonnent ton œuvre. Enfin, cette introduction m'a semblé d'autant plus délicate que, si nos chemins se sont croisés à diverses reprises, j'éprouve quelques difficultés à partager ta fascination pour les sciences cognitives... or c'est justement ce dont nous t'avons demandé de parler ici. Aussi, pour me tirer du mauvais pas dans lequel je me suis mis, je vais tâcher de m'en tenir à une présentation stroboscopique de ton itinéraire, et schématique à l'excès.

"La vie, as-tu écrit, est un jardin aux sentiers qui bifurquent" et, de fait, fidèle à la nouvelle de Borgès, ton parcours n'a cessé d'emprunter des chemins de traverse.

Entré à l'École Polytechnique "sans l'avoir choisi", au corps des Mines, pour avoir trop bien fait ce que l'on te demandait de faire, tu aurais partagé la piaule d'un ingénieur des Mines qui a lui-même partagé sa carrière entre l'atome et l'agronomie. Polytechnicien désenchanté par le monde qui séparait tes croyances positivistes et l'absurde réalité des tâches administratives qui t'attendaient, tu cherches une issue dans la formalisation économétrique. Mais, déçu par le formalisme de l'économie mathématique, te voici, selon ton expression "polytechnicien défroqué". Crise sceptique, ou relativiste, grâce à laquelle tu deviens un "intellectuel". Mais, dans la confusion des idéologies alternatives qui t'attirent, tu trouves une voie : Ivan Illich et la contre-productivité de la société industrielle. Tu mets alors ton savoir-faire d'ingénieur au service de la cause, et cela a donné quelques textes fort réjouissants. Ils prouvent, chiffres en main, que passés certains

seuils critiques, les institutions de notre société produisent le contraire de ce que nous attendons d'elles, et de ce pourquoi elles ont été mises en place. La médecine rend malade, l'école rend idiot, les transports immobilisent et la communication ne produit que mutisme et surdité. Santé, culture, savoir, liberté de mouvements et de paroles ne peuvent reposer que sur l'autonomie des individus ou de petits groupes sociaux. Certes, il faut des institutions hétéronomes, sinon les hommes ne pourraient vivre ensemble. Mais leur rôle doit se limiter à favoriser les conditions matérielles et symboliques de l'autonomie. Dans la société industrielle, ces institutions hétéronomes détruisent l'autonomie. Il faut alors, pour que tienne la société, mettre en place de nouvelles institutions qui, à leur tour, détruiront l'autonomie. Je paraphrase ce que nous aurions pu dire, il y a tant d'années déjà.

1976, à Cuervanica, au cours de la dernière des réunions mexicaines qu'organisait Illich, plus semblables à ces aimables compagnies dans lesquelles Montaigne aimait à débattre qu'à des colloques, des séminaires et des *workshops*, Jean-Pierre Dupuy rencontre Heinz Von Foerster, un des pionniers de la cybernétique. Celui-ci le convainc que, dans cet irrationnel apparent qu'est la contre-productivité de la société industrielle, il y a du rationnel. La passion formalisatrice du "polytechnicien défroqué" se réveille. Ayant lu, puis rencontré, sur les conseils de Von Foerster, Henri Atlan et Francisco Varela, Jean-Pierre se persuade que les théories de l'auto-organisation (ou de l'*autopoiesis*), qu'ils ont conçues pour modéliser l'organisation du vivant, ont une pertinence pour penser le social. C'est comme si la contre-productivité de la société industrielle avait attendu, pour être appréhendée, ces outils que l'on avait forgés pour les sciences naturelles.

Nouvelle bifurcation, nouvelle voie – provisoire – tracée. "Entre l'autonomie des gens et des petits groupes, que défend Illich, l'autonomie du social que Castoriadis pose comme l'exigence historique de notre temps, l'autonomie du seul

sujet social historique que Girard reconnaisse – la violence –, l'autonomie du vivant que cernent Atlan et Varela : n'y a-t-il pas des passerelles à établir, des chemins à frayer?"¹ . Jean-Pierre Dupuy va donc chercher le passage entre les théories de l'auto-organisation et les idéologies de l'autonomie. Peut-être est-ce celui que Michel Serres a indiqué, dans son *Passagè du Nord-Ouest* ?

En 1981, Jean-Pierre organise un colloque à Cerisy² : "*L'auto-organisation. De la physique au politique*". Ce fut un événement important de l'épistémologie contemporaine et l'on ne saurait trop inviter à en lire les actes. La même année, avec Jean-Marie Domenach et Paul Dumouchel, il crée le Centre de recherches sur l'épistémologie et l'autonomie, qu'il va diriger jusqu'à ces derniers mois. Retour à l'École Polytechnique, qui accueille le CREA. Le CREA, la création, l'ordre qui naît du désordre. La même année (peut-être un peu plus tôt, peut-être un peu plus tard), Michel Serres publie *Genèse*, où il dit comment l'harmonie naît du bruit.

Jusqu'en 1987, le CREA va poursuivre des recherches dans deux domaines que Jean-Pierre va s'efforcer, non sans mal, de mettre en synergie : les unes portent sur l'auto-organisation du vivant, les autres sur les conceptions de l'autonomie dans la philosophie sociale et politique. S'il participe aux deux programmes, cherchant toujours les passerelles (ou le passage), c'est lors d'un séminaire organisé par le second (sur "les modèles formels de la philosophie sociale et politique"), que Jean-Pierre découvre et étudie John Rawls. Un nouveau chemin de traverse s'ouvre à lui, qui le conduira à s'orienter sur la voie de la philosophie morale.

1 . J.P. Dupuy : *Ordres et désordres. Enquête sur un nouveau paradigme*, Seuil, 1982 ; p. 17.

2 . Il se poursuit en septembre à Stanford, où J.P. Dupuy enseigne la philosophie sociale et politique depuis 1984. J.P. Dupuy, P. Dumouchel (ed), *L'auto-organisation. De la physique au politique* (Actes du colloque de Cerisy), Seuil, 1983.

En 1987, pour des raisons parmi lesquelles je ne suis pas parvenu à démêler les nécessités des hasards, ni les opportunités des opportunismes, le CREA change de nom : voici le sigle rebaptisé en Centre de recherche en épistémologie appliquée. Il change aussi, du moins partiellement, de substance. C'est à la même époque que les travaux sur l'auto-organisation du CREA attirent des chercheurs – souvent formés dans des universités américaines – spécialisés en sciences cognitives. Ils y sont accueillis, au point que le CREA est aujourd'hui l'un des principaux centres de recherches cognitivistes (peut-être est-il même le principal) en France.

Tu te dis convaincu, Jean-Pierre, qu'il s'agit là d'une évolution logique, puisqu'après tout les théories de l'auto-organisation ne seraient qu'une région des sciences cognitives et qu'elles se sont développées de concert. C'est bien ce dont je doute. L'association du holisme méthodologique des premières et du réductionnisme ontologique des secondes est-elle simplement durable ? Comment, en outre, un philosophe dualiste comme toi, qui tient au monde des idées, peut-il adhérer au monisme cognitiviste, à son matérialisme radical ? C'est pourquoi je me demande si l'intérêt croissant que tu portes à la philosophie morale ne sera pas un attracteur qui t'éloignera des sciences cognitives, ou qui du moins rendra les passerelles plus difficiles à établir, plus rares donc et plus chancelantes. En d'autres termes, je me demande si tu n'es pas à une nouvelle croisée des chemins, si cette conférence s'intègre dans le bilan d'un compagnonnage déjà révolu, ou si elle est l'ébauche d'un programme nouveau.

Voilà pour ce que je sais de l'itinéraire. Mais que dire de Jean-Pierre Dupuy ? Quelles continuités établir du polytechnicien au spécialiste de philosophie morale, de l'économètre au philosophe impliqué dans le développement des sciences cognitives ? Il est possible d'en déceler trois : une ambition modulée par une inquiétude, une façon de faire et une manière d'être.

Pour l'ambition, je lui laisserai la parole. Dans une petite notice bibliographique qu'il m'a remise hier, Jean-Pierre l'expose explicitement. Sous les différents habits qu'il adopte, il *"s'efforce désespérément d'établir des ponts, ou du moins de jeter des passerelles, entre des types de pensée qui font tout pour s'éviter : la philosophie continentale et la philosophie analytique, la philosophie et les sciences de l'homme, les sciences sociales et les sciences cognitives"*. S'il poursuit, en écrivant que *"son engagement philosophique résolu devrait aider à faire oublier ses origines impures : ancien élève de l'École Polytechnique, ingénieur général des Mines (...)"*, ce n'est, à mon avis, que coquetterie bien compréhensible. Non seulement Jean-Pierre est resté, sinon fidèle du moins marqué par ses origines (d'où son intérêt pour la critique philosophique des fondements de la théorie du choix rationnel), mais tout impures qu'elles puissent paraître à ses yeux, elles l'ont bien aidé à construire les modèles formels de ses passerelles.

Voilà pour ce qui est de l'ambition ; pour l'inquiétude, je lui laisserai aussi la parole. Ce sont les dernières phrases de l'introduction à *Ordre et désordres*. *"L'époque est déjà lointaine, écrivait-il ainsi en 1981, où nous ne pouvions consacrer que nos moments perdus à des travaux qui sont à présent notre raison d'être. Le seuil critique illichien, au-delà duquel une institution devient le principal obstacle aux buts qu'elle est censée servir, est-il dépassé ?"* Combien, parmi nous, partagent cette crainte, ou celle proche qui nous prend lorsque, sachant que toute production (y compris celle des connaissances) subit la loi des rendements décroissants, nous nous apercevons que les travaux que nous avons accumulés nous placent en haut de courbe, là où la sigmoïde s'infléchit vers l'horizon ?

Une façon de faire ? Toutes les aventures intellectuelles dans lesquelles Jean-Pierre Dupuy s'est engagé sont faites de

rencontres : Illich, Girard (dont j'ai trop peu parlé), Von Foerster, Atlan, Varela, Rawls. Rencontres importantes par les traverses qu'elles lui ont fait prendre, dettes explicites d'une pensée qui s'est construite avec les autres (parfois aussi contre eux), dans une sorte de commerce incessant des idées.

Une manière d'être ? Une certaine permanence dans la diversité des engagements ? Qu'il ait été inspiré par Illich, Girard, Castoriadis, ou Rawls, qu'il se soit "coltiné" aux théories de l'auto-organisation, aux sciences cognitives ou aux théories du choix rationnel, j'ai toujours été frappé par la jubilation avec laquelle Jean-Pierre fait usage de sa raison. Enfin, si son parcours l'a conduit des Landes à l'École Polytechnique, de Paris au Mexique, du Mexique au Brésil, du Brésil à Stanford ou bien à Cerisy, Jean-Pierre est originaire de Geaune et y reste attaché. Geaune, la capitale du Tursan (il y a du Tursan blanc et du Tursan rouge, vins aimables et sans prétention), à soixante kilomètres de chez moi. Et c'est peut-être parce qu'il a cet esprit de terroir du Sud-Ouest (si manifeste dans sa cuisine), un esprit de terroir curieux et ouvert sur le monde – celui de la géographie et celui des idées – que son itinéraire peut paraître erratique à ceux qui, quand bien même auraient-ils passé leur vie dans la capitale et beaucoup voyagé, ne sont jamais sortis de leur trou.

Raphaël Larrère
Directeur de recherche INRA

Les savants croient-ils en leurs théories ?

Une lecture philosophique de l'histoire des sciences cognitives

*Il y a des gens qui peuvent croire en tout ce qu'ils veulent ;
ce sont d'heureuses créatures*

G. C. Lichtenberg

Introduction

"L'homme est né libre" mais partout on le voit soumis à des mécanismes. Mécanismes mentaux : les habitudes, les routines, la culture lui dictent des croyances, des attitudes et des conduites stéréotypées ; à l'extrême, névroses et psychoses l'assujettissent à des automatismes de répétition. Mécanismes sociaux : de la vendetta qui règle implacablement le cycle des actes vengeurs aux lois d'airain de l'économie marchande, les hommes se sont souvent pensés comme des êtres asservis à des processus qui les commandent depuis une extériorité fondatrice. Cette extériorité a d'abord été divinisée : quand Héraclès cède à une furie homicide et massacre tous les siens, c'est la déesse de la Rage qui l'a visité. A l'âge positif, la science a pris le relais : lorsque la Bourse entre dans une crise panique, c'est le mécanisme des anticipations qui s'est déréglé.

Les sciences de l'homme et de la société ont, tout au long de leur histoire, lorsqu'elles ne penchaient pas vers une forme ou une autre d'organicisme, pris au sérieux cette métaphore de la machine humaine et sociale. C'est le cas de la première d'entre elles, chronologiquement parlant : l'économie politique. Lorsqu'au XVIII^e siècle, en plein cœur des Lumières écossaises, Adam Smith fonde cette discipline, c'est à un immense système auto-organisé et autorégulé qu'il compare la société. A un autre bout du spectre, on sait l'usage et peut-être l'abus que le structuralisme et le post-structuralisme ont

fait de la notion de machine. On se souvient de Lacan : "*La question de savoir si (la machine) est humaine ou pas est évidemment toute tranchée – elle ne l'est pas. Seulement, il s'agit aussi de savoir si l'humain, dans le sens où vous l'entendez, est si humain que ça*" ; et de conclure : "*le monde symbolique, c'est le monde de la machine*³."

C'est dire que l'apparition et le développement rapide de ce nouvel ensemble de disciplines que l'on recouvre sous le nom de sciences cognitives ne constituent pas la nouveauté radicale que tant leurs promoteurs que leurs détracteurs se plaisent à souligner. Ce n'est pas en mécanisant l'humain qu'elles innovent. Ce qui fait leur spécificité, c'est le type de machine qu'elles mobilisent : il ne s'agit plus de l'horlogerie de la mécanique rationnelle à laquelle a recours Adam Smith, ni des machines hydrauliques et thermodynamiques d'un Freud. Le concept de machine qui est à la base des sciences cognitives est celui qui naît dans les années 30 avec les travaux de Gödel, de Church et de Turing. C'est l'*algorithme*, ou "machine récursive", c'est-à-dire une procédure logique qui enchaîne des opérations selon des règles fixes. Mécanisme de calcul, donc, ce qui fait que l'on dit parfois que les sciences cognitives réalisent le programme que le fondateur de la philosophie moderne, Thomas Hobbes, avait annoncé en formulant au milieu du XVII^e siècle son fameux : "*Penser, c'est calculer*". Cette référence témoigne, là encore, que la nouvelle science de l'esprit, qui entend étendre la démarche physico-mathématique à un domaine qui était jusque-là l'apanage de la philosophie et de la psychologie, plonge ses racines dans l'histoire longue de l'aventure occidentale de la connaissance.

Bref, les sciences de la cognition nouvelles venues et les déjà vénérables sciences de l'homme et du social ont et devraient avoir beaucoup de choses à se dire, dans un esprit

3. *Séminaire*, 1954-1955.

de réciprocité et de respect mutuel. Ce n'est pas un hasard si deux des fondateurs des sciences cognitives, Herbert Simon, le pionnier de l'Intelligence artificielle, et Friedrich Hayek, l'un des inventeurs du concept d'auto-organisation, sont prix Nobel d'économie. Prenons deux exemples. La psychologie cognitive fait aujourd'hui grand cas de ce qu'elle appelle la capacité "méta-représentationnelle" des êtres humains, c'est-à-dire leur faculté de se mettre à la place d'autrui, de lui attribuer des états mentaux du type croyances, intentions, désirs, etc., de *simuler* ses raisonnements, de voir le monde à travers ses yeux. Sans une telle capacité, inscrite dans un substrat neurologique, la science cognitive découvre qu'il n'y aurait pas de communication humaine possible, donc de culture, ni de morale – car le premier geste de celle-ci est bien de se porter au-devant de l'autre et de forger des représentations au sujet de ses représentations. Or, cette capacité est une vieille connaissance des sciences sociales individualistes qui naissent au sein de ces Lumières écossaises déjà évoquées, par exemple dans la philosophie d'un David Hume – lequel pouvait écrire dans son *Traité* : "*Les esprits des hommes sont des miroirs les uns pour les autres*". La théorie de la *spéculation* financière d'un Keynes, et ses avatars récents, par exemple, découlent de cette intuition initiale.

Le deuxième exemple est à dessein très éloigné du premier. On peut défendre l'idée que les sciences cognitives participent, peut-être sans le savoir, du mouvement général de "déconstruction" de la figure classique du sujet que nous ont léguée Descartes et Leibniz. De leur ancêtre cybernétique aux développements actuels, elles nous présentent en effet le sujet individuel comme un *quasi-sujet*, c'est-à-dire comme un collectif manifestant les propriétés de la subjectivité. Lorsque je pense, je me souviens, je désire, je crois, je décide, etc., ce n'est pas un fantôme dans la machine cérébrale, un homunculus caché, qui est le sujet de ces prédicats – c'est la machine elle-même, sous la forme, par exemple, d'un réseau de neurones. Il n'y a

pas de fantôme dans la machine, il n'y a pas de centre de la subjectivité, comme y insiste un Dan Dennett. Les attributs de la subjectivité sont des effets émergents produits par le fonctionnement spontané, "auto-organisé", d'une organisation complexe en forme de réseau. Les spécialistes de sciences cognitives qui défendent ce genre de thèses utilisent des expressions étranges pour les dire. Ils parlent de "*selfless selves*", d'"egos non-egoïques", de "processus sans sujet". Ce sont là des formules que des structuralistes comme Lévi-Strauss, Lacan ou Althusser, par exemple, ont employées ou auraient pu employer.

Oui, il y a matière à débats, confrontations, défis, enrichissement mutuel. Le danger, si danger il y a, tient à ceci. Pour des raisons qui tiennent aux circonstances et aux jeux de pouvoir qui accompagnent toute montée en puissance d'un nouveau champ disciplinaire, l'institutionnalisation des sciences cognitives s'est faite, dans notre pays, sous le leadership alterné des neurosciences et de l'intelligence artificielle, c'est-à-dire des biologistes et des ingénieurs, chacune de ces deux forces tentant d'arracher le pouvoir à l'autre. La lutte entre rivaux est souvent plus meurtrière pour le tiers auquel nul ne prête attention et que l'on piétine en passant que pour les rivaux eux-mêmes. Le tiers, ici, c'est la science de l'homme. Les sciences cognitives ont le vent en poupe ; on leur accorde crédits, postes, honneurs – en partie parce qu'on en attend des retombées techniques faramineuses. Dans leur ardeur de néophytes, certains de leurs champions projettent de reconstruire les sciences de l'homme en faisant table rase de tout ce qui s'est pensé jusqu'à eux. Il en résulte parfois des constructions gratuites où l'inculture le dispute à la naïveté. Il n'y a là rien de nécessaire. On pourra s'en convaincre en revenant sur certains des grands moments de l'aventure intellectuelle exaltante que constitue l'essor des sciences cognitives et en examinant quelques questions ouvertes par le défi que lancent ces disciplines aux sciences de l'homme constituées.

Première partie

Les savants croient-ils en leurs théories ?

Pour introduire mon sujet, je vais recourir à une étude de cas personnelle, qui me permettra de faire travailler devant vous l'une des catégories centrales de la philosophie de l'esprit (*philosophy of mind*), cette branche de la philosophie dite "analytique" qui encadre et informe les recherches menées en sciences cognitives. Je fais référence à la catégorie de *croyance*, dont les philosophes de la cognition nous disent qu'elle est un état mental et, plus précisément, une attitude propositionnelle, c'est-à-dire une attitude (représentée ici par le verbe "croire") dont le contenu est une proposition relative au monde.

Je m'intéresse ici à la relation très particulière qui lie croyance et vérité. Il est *constitutif* d'une croyance, nous dit-on, qu'elle se représente à elle-même son contenu comme vrai et, plus encore, qu'elle ne se juge vraie, en tant que croyance, que pour autant que son contenu est vrai. Il existe, certes, des croyances fausses, mais elles ne sont croyances que si la condition précédente est respectée. Les philosophes de l'esprit traduisent cette condition en disant que la croyance "vise" (*aims at*) la vérité.

Une façon de voir ce qu'il y a ici de singulier est de réfléchir au fameux "paradoxe de Moore". Moore fait remarquer qu'il serait absurde d'énoncer une phrase telle que :

p est le cas, mais je crois que *p* n'est pas le cas

[Par exemple : "Dieu n'existe pas, mais je crois que Dieu existe".]

Il n'y aurait, par contraste, aucune difficulté à affirmer :

p est le cas, mais j'imagine que *p* n'est pas le cas,

ou bien :

p est le cas, mais je fais semblant de croire que *p* n'est pas le cas.

Aucun problème non plus s'il s'agit de dire :

p est le cas, mais jusqu'ici j'ai toujours pensé que p n'était pas le cas,

ou encore :

p est le cas, mais Jean croit que p n'est pas le cas.

Ceci montre bien que le problème n'est pas dû à la difficulté de concevoir la possibilité de croyances fausses. Comme l'écrit Wittgenstein, lorsque nous affirmons *p*, nous affirmons que nous croyons que *p*, et réciproquement. Et cependant, ajoute le même Wittgenstein, pour qui c'est là l'essence du paradoxe de Moore, l'hypothèse que *p* est le cas n'est pas identique à l'hypothèse que nous croyons que *p* est le cas⁴. Nous pouvons fort bien dire, en effet :

Supposons que p soit le cas, mais que je croie que p n'est pas le cas.

La croyance vise donc la vérité. Mais, comme le souligne Peter Railton⁵, cette manière de présenter les choses est trompeuse, en ce qu'elle suggère une visée de type *instrumental*, exprimable en termes de moyens et de fins. En tant qu'organismes naturels, nous chercherions à maximiser globalement le vrai comme rapport d'adéquation au monde qui nous entoure – de même qu'en tant qu'êtres moraux, nous chercherions à maximiser globalement le bien –, et la formation de croyances au sujet du monde serait un moyen efficace de réaliser cette adéquation. Or tel ne peut être le cas et il importe de comprendre pourquoi.

A un moment de sa vie, le philosophe Donald Davidson s'est vu devenir chauve : trop d'indices allaient dans le même sens, de la tonsure qui se forme sur le haut de la tête et que

4. Ludwig Wittgenstein, *Philosophical Investigations*, New York, Macmillan, 1953.

5. P. Railton, "Truth, Reason, and the Regulation of Belief" in E. Villanueva (ed.), *Truth and Rationality*, Philosophical Issues, 5, Atascadero, Ca., Ridgeview Publishing Company, 1994.

les coups de peigne arrivent de plus en plus mal à masquer jusqu'aux cheveux que l'on trouve de plus en plus nombreux sur la brosse chaque matin. La blessure d'amour-propre était cruelle et Davidson s'est arrangé, comme nous savons tous le faire, pour se dissimuler la gênante vérité. De sorte que l'on pouvait décrire son état mental par les deux propositions suivantes, aux contenus contradictoires :

- (1) Davidson croit qu'il devient chauve ;
- (2) Davidson croit que la calvitie ne le menace pas.

Davidson avait de "bonnes raisons" d'avoir la croyance exprimée par (2) : cette croyance était un moyen efficace de maximiser son bien-être ou ses satisfactions d'amour-propre. On admettra, cependant, que ces raisons *prudentielles* ou *instrumentales* ne sont pas à mettre sur le même plan que les raisons *épistémiques*⁶ qui conduisent à la formation de la croyance exprimée par (1) ; mais aussi que *seules* ces raisons épistémiques ont vocation à former la croyance et que les raisons prudentielles sont en définitive de mauvaises raisons ; donc, que la croyance exprimée par (2) est une croyance fausse et irrationnelle. L'état mental caractérisé par la conjonction de (1) et de (2) est un état de *self-deception*, ou

6. Il convient en principe de distinguer "épistémique" et "épistémologique". Alors que le second terme renvoie à ce qui relève de l'épistémologie, cette discipline philosophique qui interroge les conditions de validité du discours *scientifique*, le premier désigne ce qui a trait à la *connaissance* en général, à ses mécanismes, à sa logique ou à ses fondements. Le terme "*epistemology*", en anglais, et l'adjectif qui lui correspond, "*epistemological*", recouvrent à peu près ce que le français entend par "épistémique". L'influence qu'exerce aujourd'hui la pensée anglo-américaine sur ces questions explique que la distinction entre "épistémique" et "épistémologique" tend à se perdre. Cette confusion d'ordre sémantique est bien en phase avec une tendance que nous allons analyser, qui conduit à penser les processus de la connaissance en général selon le modèle exclusif de la connaissance scientifique. L'essor des sciences cognitives est contemporain de cette tendance.

mensonge à soi-même, ou "mauvaise foi" (au sens sartrien du terme), dont Davidson, précisément, est le grand théoricien⁷.

Supposons, maintenant, que, dans des circonstances particulières, le fait d'avoir aujourd'hui telle croyance que je tiens pour fausse – disons la croyance que **p** –, conduite à maximiser l'étendue de mes croyances vraies à long terme. J'ai maintenant une "bonne raison", prudentielle ou instrumentale, d'avoir la croyance que **p**, mais cette raison est, cette fois, elle aussi épistémique. C'est, dans ce cas, à l'intérieur de la seule logique de la connaissance que le conflit existe, entre des raisons épistémiques instrumentales de croire **p** et des raisons épistémiques reposant sur les données qui me sont accessibles – je les dirai "évidentielles"⁸ – de croire **non-p**. Ce que le paradoxe de Moore nous apprend au sujet des croyances nous interdit de tenir les raisons épistémiques instrumentales pour pertinentes face aux raisons épistémiques évidentielles. Ce n'est qu'en arrivant à produire en soi un état irrationnel de *self-deception* que l'on pourrait, dans ces circonstances hypothétiques, maximiser ses croyances vraies, puisqu'il faudrait, tout à la fois, croire une chose et croire son contraire.

Certes, dira-t-on, mais il s'agit là d'un pur cas d'école, d'une invention de philosophes dépourvue de portée. Erreur : ce que ce cas formalise est en vérité le problème central de la philosophie des sciences (ou "épistémologie", au sens français du terme). Selon certains auteurs, il est possible de "dépasser" l'opposition entre Popper et Kuhn de la manière suivante⁹. D'un côté, un falsificationnisme "naïf" mettant en scène des savants empressés de tirer à boulets rouges sur leurs propres hardies conjectures, afin que seules les

7. D. Davidson, "Deception and Division", in J. Elster (ed.), *The Multiple Self*, Cambridge, UK, Cambridge University Press, 1986. Cf. J.P. Dupuy (ed.), *Self-Deception and Paradoxes of Rationality*, CSLI Publications, Stanford University, 1998.

8. Je me résous ici à cet anglicisme, de "evidence", qui signifie indice, manifestation, symptôme.

meilleures résistent à ces assauts cruels. Désintéressés, dégagés de tout préjugé ou idéologie, ces êtres rationnels cherchent la vérité sur la base des seules données de l'observation ou de l'expérience. De l'autre, une vision kuhnienne des choses selon laquelle les communautés scientifiques ont des mœurs de type villageois préindustriel, soucieuses avant tout de se plier aux normes paradigmatiques, précautionneuses, dogmatiques, attentives au qu'en-dira-t-on ; des communautés où l'attachement aux noyaux durs des programmes de recherche relève davantage de la foi religieuse que de l'adhésion rationnelle. Ou bien, pis encore, avec la version *hard* de la sociologie et de l'histoire des sciences, la vision d'un champ de bataille où l'issue des conflits dépend des seuls rapports de force, où tous les coups sont permis dès lors qu'ils rapportent au "paradigme", et surtout à ses supporters, prestige, pouvoir, postes et financements. Cette seconde représentation a certes le mérite de ne pas être naïve, mais il y a une question de base qu'elle ne songe même pas à poser, tant elle est occupée à déconstruire l'image d'une science progressant rationnellement : comment, de cette foire d'empoigne, peut-il surgir quelque chose de l'ordre de cette objectivité et de cette opérationnalité dont il est difficile de dire, sans mauvaise foi, qu'elles ne caractérisent pas l'aventure techno-scientifique ?

Il est une réponse à cette question dont le principe ne devrait pas surprendre les économistes : une explication en termes de "main invisible" – ou plutôt d'"esprit invisible", pour reprendre la suggestion de Railton¹⁰ – pourrait bien se révéler valide. De même que, sur le marché, les vices privés engendrent les bienfaits publics, il se pourrait que des

9. Cf. P. Railton, "Truth, Reason, and the Regulation of Belief", loc. cit. Des spécialistes de l'œuvre de Popper, tels qu'Alain Boyer et David Miller, font remarquer que ce prétendu dépassement est en fait déjà chez Popper.

10. "Truth, Reason, and the Regulation of Belief", loc. cit.

scientifiques aux antipodes d'une attitude poppérienne engendrent, sans le savoir ni le vouloir, un monde poppérien. Il n'est pas très difficile d'imaginer des mécanismes qui puissent rendre compte de cet "effet pervers". La rivalité de communautés dont chacune s'accroche dogmatiquement à son paradigme, la course aux honneurs et à la reconnaissance, garantissent qu'un grand nombre de pistes seront ouvertes et poursuivies avec ténacité, ce qui permettra en principe aux meilleures d'apparaître et d'être sélectionnées ; plus vraisemblablement, en tout cas, que si des savants minés par le doute soumettaient en permanence leurs hypothèses au verdict des données de l'expérience et de l'observation. On pourrait détailler, affiner l'analyse, mais la question qui ici m'intéresse est celle des conditions de possibilité *épistémiques* d'un tel fonctionnement de la science. Que se passerait-il si les acteurs eux-mêmes se persuadaient que c'est ainsi que ce fonctionnement est orienté vers l'objectivité et la vérité ? Si la science fonctionne de cette manière, n'est-il d'ailleurs pas inévitable que ses acteurs, tôt ou tard, en prennent conscience ?

Nous serions alors exactement dans le cas de figure que nous avons envisagé. Les tenants du paradigme **p** sauraient que la raison majeure qu'ils ont de croire à son "noyau dur" n'est pas de type épistémologique évidentiel – il y a peu de chances qu'il résiste à la falsification –, mais bien de type épistémologique instrumental : ce n'est que si chacun s'accroche à son paradigme que, par les mécanismes suggérés ci-dessus, la vérité pourra émerger, par effet de système en quelque sorte. Peut-on continuer à croire en **p** dans ces conditions ? Ce que nous a appris le paradoxe de Moore nous incite à répondre par la négative. Or, si les tenants d'un paradigme cessent d'y croire, la mécanique globale d'atteinte de la vérité et de l'objectivité s'enraye.

Pour résoudre des difficultés de ce type – en particulier, celles qui ont trait aux paradoxes de la *self-deception* –, les