

Daniel C. Dennett

*La Conscience
expliquée*



EDITIONS
ODILE JACOB

PHILOSOPHIE

LA CONSCIENCE EXPLIQUÉE

Ouvrages du même auteur
traduits en français

Vues de l'esprit (avec Douglas Hofstadter), InterÉditions,
1987.

*La Stratégie de l'interprète : le sens commun et l'univers
quotidien*, trad. P. Engel, Gallimard, 1990.

DANIEL C. DENNETT

LA CONSCIENCE EXPLIQUÉE

*Traduit de l'anglais (États-Unis) par Pascal Engel
Illustrations de Paul Wiener*

Ouvrage traduit avec le concours
du Centre National du Livre



Pour Nick, Marcel et Ray

L'édition originale en langue anglaise
de cet ouvrage a été publiée en 1991
par Little, Brown and Company sous le titre :
Consciousness Explained
© Daniel C. Dennett, 1991

Pour la traduction française :
© ÉDITIONS ODILE JACOB, OCTOBRE 1993
15, RUE SOUFFLOT, 75005 PARIS
ISBN : 2-7381-0216-6

La loi du 11 mars 1957 interdit les copies ou reproductions destinées à une utilisation collective. Toute représentation ou reproduction intégrale ou partielle faite par quelque procédé que ce soit, sans le consentement de l'auteur ou de ses ayants cause, est illicite et constitue une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

Note du traducteur

Il est naturel qu'un livre qui entend s'opposer à la tradition cartésienne et phénoménologique en philosophie conduise le traducteur à adopter ce qui peut apparaître aux yeux du lecteur français comme des solécismes philosophiques. C'est ainsi que j'ai gardé l'usage de l'auteur et de nombreux philosophes de langue anglaise, qui consiste à employer le mot « phénoménologie » au sens objectif, pour désigner l'univers des expériences et des qualités phénoménales perçues d'un sujet, plutôt que l'étude de ces mêmes expériences et qualités, pour lesquelles Daniel Dennett a forgé un terme, celui d'*hétérophénoménologie*.

J'ai aussi adopté les choix suivants. Pour *intentional stance*, j'ai ici adopté « posture intentionnelle » (cf. note 4, p. 131). *Multiple drafts* a été traduit par « versions multiples » et non par « brouillons multiples », car certaines versions peuvent être plus élaborées que des brouillons. *Design* pose un problème classique, en raison de la variété des sens, et ne pouvait recevoir une traduction unifiée. J'ai traduit, selon les cas par « conception », « plan », ou « organisation naturelle » dans les contextes où il est question de l'évolution, par « construction » dans d'autres cas. J'ai aussi suivi un usage répandu qui consiste à garder l'anglais *fitness*. *Report* a été traduit par « compte rendu » quand il s'agit de la forme nominale, et par « rapporter » quand il s'agit de la forme verbale (*to report*). *Central Meaner* a été traduit par « Signifieur Central ». J'ai habituellement traduit *input* et *output* par « entrée » et « sortie », sauf dans certains cas où j'ai cru bon de garder l'anglais. De même pour *hardware* et *software*, « matériel »

et « logiciel ». J'ai emprunté au traducteur italien Lauro Colasanti sa traduction ingénieuse de moi *selfy* et *sexy* par *egoici* et *erotici* : « égoïque » et « érotique ».

Je remercie bien vivement Daniel Dennett d'avoir, comme par le passé, accepté de relire le manuscrit de cette traduction et de m'avoir suggéré d'utiles corrections, ainsi que Joëlle Proust et Claudine Tiercelin. Il va de soi que je suis responsable des erreurs qui pourraient demeurer.

Préface

C'est en lisant les *Méditations* de Descartes, durant ma première année d'université, que j'ai commencé à être fasciné par le problème des relations entre l'esprit et le corps. Un mystère se présentait à moi. Comment mes pensées et mes sentiments pouvaient-ils trouver leur place dans le même monde que celui des cellules nerveuses et des molécules composant mon cerveau ? Aujourd'hui, après avoir réfléchi, parlé, écrit sur ce mystère pendant trente ans, je pense avoir fait quelques progrès. Je crois pouvoir esquisser les grandes lignes d'une solution et proposer une théorie de la conscience qui donne des réponses (ou qui montre comment trouver les réponses) aux questions qui ont suscité la perplexité des philosophes aussi bien que des savants et des profanes. Pour ce faire, j'ai été beaucoup aidé. J'ai eu la chance d'apprendre, de façon informelle, sans fatigue, et imperturbablement, ce que m'ont enseigné quelques penseurs merveilleux, que vous rencontrerez dans les pages qui suivent. Car l'histoire que je vais raconter n'est pas celle d'une cogitation solitaire, c'est une odyssée qui fait voyager à travers de nombreux domaines. Les solutions sont le fruit de dialogues et de désaccords puisque nous apprenons souvent plus des erreurs audacieuses que de prudentes équivoques. Je suis sûr qu'on peut trouver encore quantité d'erreurs dans la théorie que je présente ici. J'espère qu'elles sont audacieuses. Car si c'est le cas, elles provoqueront de meilleures réponses que les autres.

Il m'a fallu de nombreuses années pour forger les idées qu'on trouvera dans ce livre. Mais la rédaction a commencé en janvier 1990 et s'est terminée une année plus tard, grâce à la géné-

rosité de plusieurs excellentes institutions et l'aide de nombreux amis, étudiants et collègues. Le Zentrum für Interdisziplinäre Forschung à Bielefeld, le CREA à l'École polytechnique, et la Villa Serbelloni de la Fondation Rockefeller à Bellagio, m'ont fourni des conditions idéales d'écriture et de conversation durant les cinq premiers mois. Au sein de mon université, Tufts, le Center for Cognitive Studies a financé mon travail et m'a permis de présenter l'avant-dernière version à l'automne 1990 dans un séminaire auquel ont assisté les collègues et les étudiants de Tufts et des autres excellents établissements de la région de Boston. Je voudrais aussi remercier la Fondation Kapor et la Fondation Harkness pour leur soutien aux recherches du Center for Cognitive Studies.

Il y a plusieurs années, Nicholas Humphrey est venu travailler avec moi au Center for Cognitive Studies. Lui, Ray Jackendoff, Marcel Kinsbourne et moi-même avons pris l'habitude de nous rencontrer pour discuter divers aspects et problèmes liés à la conscience. Il serait difficile de trouver quatre approches de la nature de l'esprit qui soient plus différentes, mais nos discussions furent si fécondes et si encourageantes que c'est à ces bons amis que je dédie ce livre, en les remerciant de ce qu'ils m'ont appris. Deux autres collègues et amis de longue date ont eux aussi joué un grand rôle dans la formation de ma pensée. Je leur en serai éternellement reconnaissant : Kathleen Akins et Bo Dahlbom.

Je voudrais aussi remercier le groupe du ZIF à Bielefeld, en particulier Peter Bieri, Jaegwon Kim, David Rosenthal, Jay Rosenberg, Eckhard Scheerer, Bob van Gulick, Hans Flohr, et Lex van der Heiden ; le groupe du CREA à Paris, en particulier Daniel Andler, Pierre Jacob, Francesco Varela, Dan Sperber et Deirdre Wilson ; et les « princes de la conscience » qui ont rejoint Nick, Marcel, Ray et moi à la Villa Serbelloni pour une semaine intensément productive en mars : Edoardo Bisiach, Bill Calvin, Tony Marcel et Aaron Sloman. Merci aussi à Edoardo et aux autres participants de l'atelier sur les pathologies de la négligence, à Parme, en juin. Pim Levelt, Odmar Neumann, Marvin Minsky, Oliver Selfridge et Nils Nilsson m'ont aussi beaucoup aidé par leurs avis sur divers chapitres. Je voudrais aussi exprimer ma gratitude à Nils pour m'avoir fourni la photographie de Shakey, et à Paul Bach-y-Rita pour ses photos et ses avis sur les dispositifs prothésiques de vision.

Je suis reconnaissant pour une masse de critiques constructives

à tous les participants du séminaire de l'automne 1990, une classe que je n'oublierai jamais : David Hilbert, Krista Lawlor, David Joslin, Cynthia Schossberger, Luc Faucher, Steve Weinstein, Oakes Spalding, Mini Jaikumar, Leah Steinberg, Jane Anderson, Jim Beattie, Evan Thompson, Turhan Canli, Michael Antony, Martina Roepke, Beth Sangree, Ned Block, Jeff Mc Connell, Bjorn Ramberg, Phil Holcomb, Steve White, Owen Flanagan et Andrew Woodfield. Semaine après semaine, cette bande me maintint sur le grill, de la façon la plus constructive possible. Pendant la lecture finale de la dernière version, Kathleen Akins, Bo Dahlbom, Doug Hofstadter et Sue Stafford m'ont fait de nombreuses suggestions qui ont eu pour moi une valeur inestimable. Paul Weiner a transformé mes dessins grossiers en figures et diagrammes en excellents.

Kathryn Wynes et plus tard Anne Van Voorhis ont accompli un travail extraordinaire pour m'éviter, au Center for Cognitive Studies, de me disperser durant ces dernières années trépidantes. Sans leur efficacité et leur prévoyance, ce livre serait encore à des années de son achèvement. Finalement, le plus important : mon affection et mes remerciements à Susan, Peter, Andrea, Marvin, et Brandon, ma famille.

Université de Tufts, janvier 1991.

Prélude

Comment les hallucinations sont-elles possibles ?

Le cerveau dans la cuve

Supposez que des savants fous retirent votre cerveau de votre corps pendant votre sommeil et qu'ils le placent dans une cuve pourvue d'un dispositif permettant de le conserver vivant. Supposez qu'ils vous fassent croire que vous n'êtes pas seulement un cerveau dans une cuve, mais que vous êtes encore bien éveillé et en train d'accomplir avec votre corps des activités normales dans le monde réel. Cette vieille rengaine, l'histoire du cerveau dans la cuve, est l'une des expériences de pensée favorites dont se servent de nombreux philosophes. C'est une version moderne du malin génie cartésien. Descartes (1641) ¹ a en effet imaginé qu'un illusionniste cherchait à le tromper sur absolument tout, et même sur sa propre existence. Mais comme il le faisait remarquer, même un malin génie infiniment puissant ne pourrait lui faire croire qu'il existait s'il n'existait pas : *cogito, ergo sum*, (« Je pense, donc je suis »). Aujourd'hui, les philosophes sont moins intéressés par la perspective de démontrer leur propre existence comme choses pensantes. (C'est peut-être qu'ils considèrent que Descartes a résolu la question de façon satisfaisante.) Ils s'intéressent plus à ce qu'en principe, nous pouvons conclure à partir de notre propre expérience au sujet de notre nature et au sujet de la nature du monde dans lequel nous vivons (apparemment). *Pourriez-vous n'être qu'un cerveau dans une cuve ? Pourriez-vous avoir toujours été un cerveau dans une cuve ? Et si oui, pourriez-vous, à défaut de le vérifier, ne serait-ce que concevoir dans quelle situation vous vous trouvez ?*

L'idée du cerveau dans la cuve est une image utile pour explorer ces questions. Toutefois, je voudrais faire un nouvel usage de cette vieille rengaine. Je voudrais l'utiliser pour mettre en évidence certains faits curieux au sujet des hallucinations, ce qui nous mettra sur la voie d'une théorie – empirique, respectable scientifiquement – de la conscience humaine. Dans l'expérience de pensée habituelle, il est évident que nos savants devaient besogner afin de donner aux bouts de nerfs coupés de tous vos sens les stimulations nécessaires pour créer l'illusion. Pour les besoins de l'argument, les philosophes ont supposé, qu'aussi difficile que puisse être techniquement parlant la tâche, elle était « possible en principe ». Il faut se méfier de ces possibilités de principe. Il est possible aussi en principe de construire une échelle en acier inoxydable allant jusqu'à la Lune et d'écrire, par ordre alphabétique, toutes les conversations intelligibles en français composées de moins de mille mots. Mais ce n'est pas possible en fait, même de loin, et il arrive qu'une *impossibilité de fait* soit théoriquement plus intéressante qu'une *possibilité de principe*, comme nous le verrons.

Arrêtons-nous un instant, donc, pour voir à quel point la tâche qu'auraient à accomplir nos savants fous serait décourageante. Nous pouvons imaginer qu'ils partiraient de ce qui est facile avant de passer à plus difficile. Ils commenceraient avec un cerveau placé dans le coma qui convient, gardé vivant, mais privé de toute information en provenance des nerfs optiques, des nerfs auditifs, des nerfs somatosensoriels, et de toutes les entrées afférentes conduisant au cerveau. On soutient quelquefois qu'un cerveau ainsi « désafférenté » de ce genre resterait en permanence dans un état comateux, et qu'il n'aurait pas besoin de morphine pour demeurer endormi ; mais certaines données suggèrent que, même dans ces circonstances épouvantables, le cerveau pourrait encore s'éveiller spontanément. Je pense que nous pouvons supposer que si vous étiez éveillé dans un tel état, vous seriez dans une situation désespérée : vous seriez aveugle, sourd, complètement engourdi, sans le moindre sens de l'orientation de votre corps.

Puisqu'ils ne veulent pas vous horrifier, nos savants fous trouvent le moyen de vous éveiller en infusant de la musique stéréophonique (encodée astucieusement comme des influx nerveux) dans vos nerfs auditifs. Ils font aussi en sorte que les signaux nerveux qui viennent normalement de votre système vestibulaire ou de votre oreille interne vous indiquent que vous

êtes couché sur le dos, paralysé, engourdi, aveugle. Les prouesses techniques devraient pouvoir permettre ce genre de choses dans un futur proche, peut-être même dès maintenant. Nos savants fous pourraient ensuite stimuler les systèmes qui innervent votre épiderme, et lui fournir l'information qu'aurait pu produire normalement une chaleur douce et égale sur votre ventre. S'enhardissant, ils pourraient stimuler les nerfs dorsaux de votre épiderme ainsi que le ferait le picotement de grains de sable dans votre dos. « Fantastique, vous dites-vous, me voici, sur une plage, étendu sur le dos, paralysé et aveugle, en train d'écouter une musique plutôt agréable, mais probablement en danger d'attraper un coup de soleil. Comment suis-je parvenu jusqu'ici, et comment puis-je appeler à l'aide ? »

À présent supposez que nos savants, après avoir réussi tout cela, s'attaquent à un problème plus difficile : vous convaincre que vous n'êtes pas seulement un légume en train de rôtir sur une plage, mais un agent capable d'entreprendre certaines activités dans le monde. Ils décident d'éliminer progressivement la « paralysie » qui a envahi votre corps fantomatique et de vous laisser remuer votre index droit dans le sable. Ils vous permettent de faire l'expérience sensorielle correspondant au mouvement de votre doigt, en vous donnant le feed-back kinesthésique associé aux signaux volitionnels ou moteurs dans les sorties ou dans les parties efférentes de votre système nerveux. Mais ils s'arrangent aussi pour ôter à votre doigt fantôme son engourdissement et stimulent en vous l'impression que le mouvement du sable imaginaire provoquerait autour de votre doigt.

Soudain, ils rencontrent un problème : la nature de la sensation que vous procurera le sable dépendra de la manière dont vous décidez de mouvoir votre doigt. Il leur faudrait calculer le bon feed-back, l'engendrer et le composer, puis vous le présenter en temps réel. C'est une tâche computationnellement insoluble même pour l'ordinateur le plus rapide. Si nos savants fous décident de résoudre le problème lié au temps réel en précalculant et en « mettant en boîte » toutes les réponses possibles pour vous les jouer en play-back, ils échangeront seulement un problème insoluble contre un autre, car il faudra emmagasiner trop de possibilités pour que cela marche. En bref, nos savants fous seront débordés par une *explosion combinatoire* dès qu'ils vous donneront de véritables pouvoirs d'exploration du monde imaginaire² dans lequel ils veulent vous placer.

Nos savants viennent de se heurter à un mur qui nous est familier : nous en voyons l'ombre se profiler dans les stéréotypes ennuyeux que recèlent tous les jeux vidéo. Les possibilités d'action doivent être strictement – et artificiellement – limitées de sorte que la tâche de ceux qui se représentent ce monde reste faisable. Si nos savants ne peuvent rien faire de mieux que vous convaincre que vous êtes voué à jouer toute votre vie durant à Donkey Kong, ce sont vraiment des savants fous.

Ce problème technique a pourtant une forme de solution. C'est celle qu'on utilise, par exemple, pour faciliter le poids de la tâche computationnelle dans les simulateurs de vol les plus réalistes : elle consiste à utiliser des *répliques* des objets du monde simulé. On utilise un cockpit réel, et on le pousse et le tire avec des bras hydrauliques, plutôt que d'essayer de simuler tous ces effets sur le siège ou sur le pantalon du pilote qui s'entraîne. En bref, il n'existe pour vous qu'une seule manière d'enregistrer autant d'information sur un monde imaginaire à explorer et d'avoir un accès rapide à cette information : elle consiste à utiliser un monde *réel* (même minuscule, même artificiel, même en carton-pâte) pour enregistrer sa propre information ! C'est « de la triche » si vous êtes le malin génie qui prétend avoir trompé Descartes à propos de l'existence de toutes choses sans exception. Mais c'est la seule manière de s'en tirer si on dispose de ressources qui ne sont pas infinies.

Descartes a eu la sagesse de doter son malin génie de pouvoirs *infinis* de tromperie. Bien que la tâche ne soit pas, à strictement parler, infinie, la quantité d'informations que peut obtenir rapidement un être humain curieux de ce qui l'entoure est incroyablement grande. Les ingénieurs mesurent le flux d'informations en bits par seconde et parlent de la *largeur de bande* des canaux à travers lesquels s'écoule l'information. La télévision requiert une bande plus large que la radio, et la télévision à haute définition aurait une largeur de bande encore plus grande. Une télévision à haute définition sentante-touchante aurait une largeur de bande encore plus grande, et une télévision sentante-touchante *interactive* aurait une largeur de bande astronomique, parce qu'elle se brancherait en permanence sur des milliers de trajectoires légèrement différentes à travers le monde. Jetez à un sceptique une pièce douteuse, et en une seconde ou deux, la soupesant, la grattant, la faisant sonner, la goûtant, et regardant simplement comment le soleil se reflète sur sa sur-

face, il aura consommé plus de bits d'information que n'en peut organiser en une année un super-ordinateur Cray. Fabriquer une fausse pièce *réelle* est un jeu d'enfant ; forger une pièce *simulée* à partir seulement d'une structure de stimulations nerveuses dépasse les capacités de la technologie humaine, peut-être pour toujours ³.

Nous pouvons en conclure — au cas où vous seriez inquiet — que nous ne sommes pas des cerveaux placés dans des cuves. Nous pouvons aussi en déduire que les hallucinations fortes sont tout simplement impossibles ! Par « hallucination forte », j'entends une hallucination d'un objet réel apparemment concret, tridimensionnel et persistant, et non les éclairs, les distorsions géométriques, les images après-coup, les expériences fugaces de membres fantômes, et autres sensations anormales. Une hallucination forte serait, par exemple, un fantôme qui vous répondrait, qui vous permettrait de le toucher, qui résisterait avec une certaine solidité, qui projetterait une ombre et qui serait visible à partir de n'importe quel angle, de sorte que vous puissiez en faire le tour et regarder à quoi il ressemble de dos.

En gros, on peut classer les hallucinations selon leur force d'après le nombre de caractères de ce genre qu'elles manifestent. Les comptes rendus d'hallucinations *très* fortes sont rares ; si leur crédibilité semble être inversement proportionnelle à la force de l'hallucination dont ils parlent, ce n'est pas l'effet d'une coïncidence. Nous sommes — et devrions être — particulièrement sceptiques quand on nous rapporte des hallucinations très fortes parce que nous ne croyons pas aux fantômes ; nous pensons que seul un vrai fantôme pourrait produire une hallucination très forte. (C'est avant tout parce que les hallucinations rapportées par Carlos Castañeda dans *Les Enseignements de Don Juan : confessions d'un sorcier Yaqui* (1968) parlaient tant à l'imagination que les scientifiques ont d'abord suggéré que le livre, bien que ce fût une thèse de doctorat soutenue avec succès à l'UCLA, était de la fiction, pas des faits.)

Si l'on sait qu'il n'y a jamais d'hallucinations *réellement* fortes, il ne fait pas de doute que les gens font l'expérience d'hallucinations convaincantes, mettant en jeu plusieurs modalités sensorielles. Les hallucinations qui sont bien attestées dans les écrits de la psychologie clinique sont souvent fantastiques jusque dans leurs détails, bien au-delà de ce qu'est capable d'engendrer la technologie contemporaine. Comment donc un seul cerveau peut-

il réussir ce que des équipes de savants et de développeurs de logiciels considéreraient comme presque impossible à réaliser ? Si de telles expériences ne sont pas des perceptions authentiques ou véridiques d'une chose bien réelle se trouvant « en dehors » de l'esprit, elles doivent se produire entièrement à l'intérieur de l'esprit (ou du cerveau) ; elles doivent être entièrement concoctées artificiellement, mais rester assez proches de la vie réelle pour tromper l'esprit qui les façonne.

Des farceurs dans le cerveau

On suppose d'ordinaire que les hallucinations se produisent quand une autostimulation monstrueuse a lieu dans le cerveau, en particulier une stimulation d'origine totalement interne provenant de certaines parties ou de certaines strates des systèmes perceptifs du cerveau. Descartes, au XVII^e siècle, a envisagé très clairement cette hypothèse dans l'analyse qu'il donne des membres fantômes. Il s'agit d'une hallucination étonnante mais tout à fait normale par laquelle les amputés semblent ressentir non seulement la présence de leur membre amputé, mais aussi des gratouillements, des picotements et des douleurs dans leur membre absent. (Il arrive souvent que les gens qu'on vient d'amputer, juste après l'opération, ne parviennent à croire qu'on les a amputés d'une jambe ou d'un pied que lorsqu'ils *voient* qu'ils ne sont plus là, tant leur sensation de la présence continue du membre est vive et réaliste.) Descartes recourait à l'analogie du cordon de sonnette. Bien avant les sonnettes électriques, les interphones et les talkies-walkies, les grandes maisons étaient équipées de merveilleux systèmes de fils et de poulies qui permettaient d'appeler un serviteur de n'importe quelle pièce. Un petit coup tiré sur une boucle de velours pendant d'un trou dans le mur tirait sur un fil rattaché à des poulies, qui courait jusqu'à l'office, où il mettait en branle l'une des nombreuses sonnettes à fonctions spécifiques, informant le maître d'hôtel que ses services étaient requis dans la chambre du maître, au petit salon ou dans la salle de billard. Tous ces systèmes marchaient bien, mais ils donnaient parfois lieu à des farces qui ressemblaient un peu à de petites hallucinations. Il suffisait que quelque chose tire sur le fil du petit salon à un endroit quelconque du circuit pour que le maître d'hôtel déboule en hâte... et constate qu'on ne l'avait pas appelé.

De même, pensait Descartes, puisque les perceptions sont causées à l'intérieur du système nerveux par diverses chaînes complexes d'événements conduisant en dernier lieu au système de contrôle formé par l'esprit conscient, si on intervenait quelque part le long de la chaîne (à un endroit quelconque du nerf optique, par exemple, entre le globe oculaire et la conscience) et si on déclenchait juste les nerfs qu'il faut, on produirait exactement la chaîne correspondant aux effets de la perception normale et véridique de quelque chose ; cela produirait, du côté de l'esprit où sont situés les récepteurs, exactement l'effet suscité par une telle perception consciente.

Le cerveau – ou une partie du cerveau – fait par inadvertance des farces mécaniques à l'esprit. C'est ainsi que Descartes expliquait les hallucinations de membres fantômes. Bien qu'elles soient remarquablement vives, elles sont – selon notre terminologie – relativement faibles ; elles consistent à éprouver des douleurs et des démangeaisons en désordre, toutes dans une modalité sensorielle unique. Les amputés ne voient, n'entendent ni (pour autant que je sache) ne sentent l'odeur de leurs pieds fantômes. Par conséquent, une analyse assez semblable à celle de Descartes *pourrait* expliquer les membres fantômes, si on laisse pour le moment de côté l'énigme bien connue de l'interaction entre le cerveau physique et l'esprit conscient non physique. Mais même la partie purement mécanique de l'histoire de Descartes doit être fausse s'agissant des hallucinations relativement fortes ; le cerveau ne peut pas, tel un illusionniste, conserver et manipuler assez d'informations fausses pour tromper un esprit attentif. Il peut se relâcher et laisser le monde réel fournir un excès d'information *vraie*, mais s'il commence à court-circuiter ses propres nerfs (ou à tirer ses propres cordes, comme aurait dit Descartes), cela ne produit que les plus faibles des hallucinations fugitives. (De même le mauvais fonctionnement du sèche-cheveux de votre voisin peut faire apparaître de la « neige », des « parasites » ou des vrombissements et des bourdonnements sur votre écran de télévision. Mais si vous voyez une version tronquée du journal télévisé, vous *savez* qu'elle a une cause élaborée et organisée qui va bien au-delà des talents d'un sèche-cheveux.)

Il est tentant de supposer que nous avons été peut-être trop crédules au sujet des hallucinations. Peut-être les seules hallucinations bien réelles sont-elles celles qui sont atténuées, fugaces et légères, tandis que les fortes ne se produisent pas parce qu'elles

ne peuvent pas se produire ! Si l'on parcourt rapidement la littérature sur les hallucinations, on a le sentiment qu'il existe une sorte de relation inverse entre la force et la fréquence — ainsi qu'entre la force et la crédibilité. Mais cette relation nous met aussi sur la voie d'une autre théorie du mécanisme présidant à la production des hallucinations : l'un des traits endémiques des comptes rendus d'hallucinations est que la victime rapporte en général avoir éprouvé une passivité assez inhabituelle face à l'hallucination. Les hallucinés se contentent en général de se tenir debout et d'admirer. Il est typique qu'ils n'éprouvent aucun désir de confirmer, de mettre en question ou de s'interroger, et qu'ils ne cherchent pas à entrer en contact avec les apparitions. Il est probable, pour les raisons que nous venons d'indiquer, que cette passivité est une caractéristique essentielle des hallucinations ; c'est une précondition nécessaire à l'apparition d'une hallucination modérément détaillée.

La passivité, cependant, n'est qu'une des manières dont les hallucinations relativement fortes peuvent survivre. Si ces hallucinations peuvent survivre, c'est que l'illusionniste — c'est-à-dire la chose, quelle qu'elle soit, qui produit l'hallucination — peut « compter sur » le fait que la victime fera une exploration d'un certain type — et dans le cas où elle est totalement passive, aucune exploration. Tant que l'illusionniste peut prédire en détail le type d'exploration qui sera effectué, il lui suffit de préparer les conditions de l'illusion « dans les directions où regardera la victime ». Les monteurs de décors de cinéma cherchent toujours à savoir à l'avance quelle sera la localisation de la caméra et si elle ne doit pas rester statique, quelle sera sa trajectoire exacte et son angle d'orientation. Cela leur permet de préparer juste ce qu'il faut pour couvrir les perspectives qui seront effectivement filmées. (Ce n'est pas pour rien que le *cinéma vérité*⁴ fait un grand usage de la caméra vagabonde tenue à la main.) Dans la vie réelle, Potemkine utilisait le même principe pour économiser les décors de villages que la Grande Catherine devait visiter ; son itinéraire devait être cuirassé.

L'une des solutions au problème des hallucinations fortes consiste donc à supposer qu'un lien entre la victime et l'illusionniste permet à ce dernier de faire en sorte que l'illusion *dépende* des intentions et des décisions exploratoires de la victime. Il doit pouvoir anticiper. Quand l'illusionniste n'est pas capable de « lire dans l'esprit de la victime » pour obtenir cette infor-

mation, il est toujours possible dans la vie réelle pour un magicien de théâtre, par exemple, de polariser l'attention dans une direction particulière grâce à un « coup de force psychologique » subtil mais puissant. Ainsi, un magicien manipulateur de cartes dispose de nombreuses techniques pour donner à la victime l'illusion qu'elle choisit librement parmi les cartes qu'elle peut examiner sur la table, alors qu'en fait, on ne peut retourner qu'une seule carte. Pour revenir à notre expérience de pensée antérieure, si nos savants fous peuvent *forcer* le cerveau dans la cuve à avoir un certain ensemble d'intentions exploratoires particulières, ils peuvent résoudre le problème de l'explosion combinatoire en ne préparant que les matériaux anticipatoires ; le système n'aura des interactions qu'*en apparence*. De la même manière, le malin génie de Descartes peut entretenir l'illusion grâce à des pouvoirs qui ne sont pas infinis s'il peut entretenir une illusion de libre arbitre chez sa victime, dont il contrôle minutieusement toutes les investigations sur le monde imaginaire qui lui est proposé ⁵.

Mais il existe une façon encore plus économique (et plus réaliste) de produire des hallucinations dans le cerveau : il suffit de se concilier la curiosité, toujours très flottante, de la victime. On peut comprendre comment cela marche par analogie avec un jeu de société.

Un jeu de société nommé psychanalyse

Pour jouer, on demande à une personne — ce sera la dupe — de sortir et on l'informe que, pendant qu'elle sera dehors, un membre du groupe racontera un rêve récent. Tout le monde connaîtra les grandes lignes du rêve et pourra répondre — par oui ou par non — aux questions que la dupe devra poser. Le but du jeu est de deviner qui a raconté son rêve en procédant à une sorte de psychanalyse.

Une fois que la dupe a quitté la pièce, l'hôte explique au reste de l'assemblée qu'en réalité, personne ne va raconter de rêve. L'assemblée va seulement répondre aux questions de la dupe en procédant ainsi : si la dernière lettre du dernier mot de la question appartient à la première moitié de l'alphabet, on doit répondre aux questions par oui ; à toutes les autres questions, on répondra

par non. À une exception près : on ne doit pas contredire des réponses antérieures. Par exemple :

Q : Le rêve porte-t-il sur une fille ?

R : Oui.

Mais si, ensuite, notre dupe oublie et demande :

Q : Y a-t-il des personnages féminins là-dedans ?

R : Oui [En dépit du *s* final, conformément à la règle de non-contradiction ⁶].

Quand la dupe revient dans la pièce et commence à poser ses questions, elle obtient une série de réponses plus ou moins aléatoire, ou en tout cas arbitraire. Le résultat est en général assez amusant. Parfois le processus devient rapidement absurde, comme on peut l'imaginer si l'on suppose que la question initiale était « Est-ce que l'intrigue du rêve est identique trait pour trait avec celle de *Don Quichotte* ? » ou « Le rêve comporte-t-il des êtres animés ? ». La plupart du temps, on obtient une histoire bizarre et souvent obscène racontant une mésaventure ridicule, à l'amusement de tous. Quand finalement la dupe avoue que tout ce qu'il sait c'est que le rêveur doit être un individu fortement dérangé, l'assemblée rétorque avec force rires que c'est la dupe elle-même qui est l'auteur du « rêve ». Ce n'est pas totalement vrai, bien entendu. En un sens, la dupe est l'auteur en vertu des questions qu'il a eu l'idée de poser. (Personne d'*autre* n'a proposé de mettre trois gorilles dans un canot avec une nonne.) Mais d'un autre côté, le rêve n'a tout simplement pas d'auteur, et c'est là le point essentiel. Ce jeu fait apparaître un processus de production narrative, d'accumulation de détails, sans intentions de l'auteur, ni plans. C'est une illusion sans illusionniste.

La ressemblance est frappante entre la structure de ce jeu de société et toute une famille de modèles bien connus qui prétendent expliquer les systèmes perceptifs. On dit souvent que la vision humaine, par exemple, ne peut s'expliquer entièrement comme un processus « guidé par des données » ou par un processus « de bas en haut » (*bottom-up*), mais qu'elle a besoin, à ses niveaux les plus élevés du concours d'hypothèses « guidées par des attentes » (ou de quelque chose d'analogue à une procédure de vérification d'hypothèses). Parmi ces modèles, on trouve aussi celui de la perception d'« analyse par synthèse » : il suppose aussi que les perceptions sont construites selon un processus qui

fait des va-et-vient entre des attentes engendrées par les systèmes centraux d'un côté, et les confirmations (et infirmations) nées de la périphérie d'autre part (par exemple, cf. Neisser, 1967). L'idée directrice de ces théories est qu'après une certaine quantité de « prétraitement » dans les couches primaires ou périphériques du système perceptif, la perception termine sa tâche – les objets sont identifiés, reconnus, catégorisés – par des cycles de formation et de confirmation d'hypothèses. Au cours de ces cycles, nos attentes et nos intérêts du moment façonnent des hypothèses que doit confirmer ou infirmer notre système perceptif, une séquence rapide de formations d'hypothèses de ce genre produit le résultat final, le « modèle présent », mis à jour, du monde du sujet percevant. Ces analyses de la perception reposent sur tout un ensemble de considérations, à la fois biologiques et épistémologiques. On ne peut pas vraiment dire que ce modèle a été confirmé, mais les expériences qu'il a inspirées ont été réussies. Certains théoriciens ont même soutenu que la perception doit avoir une structure fondamentale de ce type.

Quelle que soit en définitive la valeur des théories de la perception comme confirmation d'hypothèses, elles renforcent une analyse simple et forte de la nature de l'hallucination. Pour qu'un système perceptif par ailleurs normal tombe dans un état hallucinatoire, il faut que la formation des hypothèses (guidée par des attentes) soit normale, tandis que la confirmation (guidée par les données) soit désordonnée, aléatoire ou arbitraire, comme dans le jeu de société. En d'autres termes, si du bruit dans le canal qui produit des données est arbitrairement amplifié dans des « confirmations » et des « infirmations » (les réponses arbitraires par oui et par non du jeu de société), les attentes présentes, les soucis, les obsessions et les préoccupations de la victime conduiront à la formation de questions et d'hypothèses dont le contenu reflétera nécessairement ces intérêts, en sorte qu'une « histoire » se déroulera dans le système perceptif sans que quiconque en soit l'auteur. Il n'est pas nécessaire de supposer que l'histoire a été écrite d'avance ni que l'information a été enregistrée ou composée dans la partie du cerveau qui joue les illusionnistes. Tout ce que nous supposons, c'est que l'illusionniste fait des confirmations arbitraires et que la dupe fournit le contenu en posant des questions.

Un lien, aussi direct que possible, est donc fait entre l'état émotionnel du sujet de l'hallucination et les contenus des hal-

lucinations produites. Les contenus des hallucinations sont en général liés aux préoccupations du sujet qui en est victime. Pour le comprendre, il n'est pas nécessaire de faire intervenir un improbable narrateur interne qui connaît ou a une théorie de la psychologie de la victime. Pourquoi, par exemple, le chasseur, le dernier jour de la saison de la chasse aux cerfs, *voit-il* un cerf, tout entier, avec ses bois et sa queue blanche, alors qu'il est en train de regarder une vache noire ou un autre chasseur en veste orange ? Parce qu'un questionneur interne lui demande obsesivement si c'est un cerf et reçoit la réponse « non » jusqu'à ce que finalement, un bruit dans le système se trouve par erreur amplifié en un « oui », engendrant ainsi des résultats catastrophiques.

Un certain nombre de données expérimentales s'accordent bien avec cette manière de comprendre les hallucinations. Par exemple, il est bien connu que les hallucinations sont le résultat normal de privations sensorielles prolongées (voir, par exemple, Vosberg, Fraser et Guehl, 1960). Cela permet d'expliquer le fait que dans la privation sensorielle, la partie du système de confirmation des hypothèses qui est guidée par des données, n'ayant aucune donnée à sa disposition, abaisse son seuil de réceptivité au bruit ; celui-ci se trouve amplifié dans des groupes de signaux arbitraires de confirmation et d'infirmité, ce qui finit par produire des hallucinations détaillées, dont le contenu n'est dû qu'à des attentes anxieuses et des confirmations aléatoires. De plus, dans la plupart des comptes rendus, on constate que les hallucinations ne s'élaborent que graduellement (sous des conditions de privation sensorielle ou par l'effet de drogues). Elles commencent par être faibles – par exemple géométriques – puis deviennent plus fortes (« objectives » ou « narratives ») ; c'est précisément cela que ce modèle nous permet de prédire (voir par exemple Siegel and West, 1975).

Pourquoi une drogue, en se diffusant dans le système nerveux, peut-elle produire des effets aussi complexes et riches en contenu ? Car il va de soi, même s'il y a des gens assez crédules pour avoir envie de le croire, que la drogue *par elle-même* ne peut pas « contenir l'histoire ». Il est improbable qu'une drogue, en agissant de manière diffuse, puisse créer ou même déclencher un système illusionniste complexe. Au contraire, il est facile de voir comment une drogue pourrait agir directement de manière à

élever, abaisser ou désorganiser arbitrairement un seuil de confirmation au sein d'un système de confirmations d'hypothèses.

Il va de soi que ce modèle d'engendrement des hallucinations inspiré du jeu de société pourrait aussi expliquer la composition des rêves. Depuis Freud, personne ne doute du fait que le contenu thématique des rêves est éminemment symptomatique des pulsions, des anxiétés et des préoccupations les plus profondes du rêveur. Mais on sait que les indices que nous fournissent les rêves sont cachés sous des couches de symbolisme et de fausses directions. Quel processus pourrait produire des histoires répondant de façon si efficace et si incessante aux préoccupations les plus profondes du rêveur, tout en habillant l'ensemble de couches de métaphores et de déplacements ? La réponse freudienne canonique consiste à supposer qu'un scénariste intérieur composerait des pièces de théâtre rêvées pour le bénéfice thérapeutique de l'ego et les ferait passer devant un censeur intérieur en déguisant habilement leur signification véritable. (Cette hypothèse pourrait s'appeler le modèle de Hamlet. Il rappelle en effet le truc d'Hamlet consistant à faire donner la pièce « Le piège à rat » juste pour Claudius. Il faut en effet un diable très malin pour imaginer un stratagème aussi subtil. Mais, si l'on doit en croire Freud, nous avons tous en nous de tels narrateurs virtuoses.) Comme nous le verrons plus tard, il n'est pas toujours impératif d'éviter les théories postulant des *homoncules* (de « petits hommes » dans le cerveau). Mais chaque fois qu'on appelle à l'aide des homoncules il vaut mieux que ce soient des fonctionnaires relativement stupides plutôt que les brillants scénaristes de Freud, qui sont supposés produire de nouvelles scènes pour chacun de nos rêves nocturnes ! Le modèle dont il est question ici élimine entièrement le scénariste et compte sur « le public » (analogue à celui qui est le « là-dedans » dans le jeu de société) pour fournir le contenu. Le public n'est pas un simple figurant, bien entendu ; mais il n'a pas besoin d'avoir une théorie de ses propres anxiétés : il doit seulement se laisser guider par elles pour poser des questions.

Il est intéressant de noter, d'ailleurs, qu'il existe une propriété du jeu de société qui n'est pas nécessaire pour un processus de production de rêves ou d'hallucination : la règle de non-contradiction. Puisque les systèmes perceptifs d'un sujet sont toujours supposés être en train d'explorer une situation présente (plutôt qu'un *fait accompli* ⁷, un récit de rêve déjà terminé et déjà raconté)

la machinerie peut considérer que les confirmations « contradictoires » qui viennent ensuite indiquent un nouveau changement dans le monde plutôt qu'une révision de l'histoire que connaissent déjà ceux qui racontent le rêve. Le fantôme était bleu quand je l'ai vu pour la dernière fois, mais il est soudain devenu vert, et ses mains sont devenues des griffes, et ainsi de suite. La versatilité des métamorphoses des objets dans les rêves et dans les hallucinations est l'un des traits les plus frappants de ces récits, et il est encore plus frappant de constater combien nous nous « inquiétons » rarement de ces métamorphoses quand nous rêvons. Ainsi, la ferme en Normandie devient soudain une banque à Naples ⁸, et le cheval que je montais est à présent un bateau, non, pardon, un hors-bord, et j'étais d'abord accompagné par ma grand-mère mais elle est devenue le Pape. Ce sont des choses qui arrivent.

Cette légèreté est précisément ce à quoi nous devons nous attendre de la part d'un enquêteur peu porté au scepticisme qui serait confronté à un échantillon de oui et de non pris au hasard. En même temps, la persistance de certains thèmes et de certains objets dans les rêves, leur refus de se métamorphoser ou de disparaître, s'expliquent très bien avec notre modèle. Si nous admettons, pour le moment, que le cerveau utilise la règle de l'alphabet et traite son information en français, nous pouvons imaginer comment ce questionnement souterrain peut créer un rêve obsessif :

Q : Cela concerne-t-il mon père ?

R : Non.

Q : Cela concerne-t-il un téléphone ?

R : Oui.

Q : Bon. Cela concerne-t-il ma mère ?

R : Non.

Q : Cela concerne-t-il mon père ?

R : Non.

Q : Est-ce à propos de mon père au téléphone ?

R : Oui.

Q : Je le savais bien que cela concernait mon père ! Eh bien, est-ce qu'il me parlait ?

R : Oui...

On peut difficilement dire que cette petite ébauche de théorie prouve quoi que ce soit quant aux hallucinations et aux rêves. Elle montre – de façon métaphorique – ce que *pourrait* être une

explication mécaniste de ces phénomènes. C'est déjà un bon début, parce que certaines personnes sont tentées d'accepter la thèse défaitiste selon laquelle la science ne peut pas expliquer « en principe » les différents mystères de l'esprit. L'ébauche en question ne pose même pas, pour le moment, le problème de notre *conscience* des rêves et des hallucinations. De plus, bien que nous ayons exorcisé un improbable homoncule, habile illusionniste/scénariste qui jouerait des tours à l'esprit, nous l'avons remplacé non seulement par des interrogateurs stupides (que l'on peut aussi « remplacer par des machines »), mais aussi par un interrogateur intelligent et inexplicé, le « public ». Si nous avons éliminé un méchant, nous n'avons même pas commencé à expliquer la victime.

Pourtant, nous avons un peu progressé. Nous avons vu comment, quand on s'intéresse à ce que requiert l'« ingénierie » d'un phénomène mental, on peut poser des questions nouvelles, qui peuvent recevoir une réponse immédiate, telles que : quels sont les modèles de l'hallucination qui permettent d'éviter l'explosion combinatoire ? Comment le contenu d'une expérience peut-il être élaboré par des processus (relativement) stupides et aveugles ? Quelles sortes de liens entre processus et systèmes peuvent expliquer les résultats de leur interaction ? Si nous devons construire une théorie scientifique de la conscience, il nous faudra poser beaucoup de questions de ce genre.

Nous avons également introduit l'une des idées centrales de la suite de ce livre. L'élément clef dans nos diverses explications du fonctionnement des hallucinations et des rêves, c'était l'idée que tout ce que le cerveau doit faire c'est de *calmer son appétit épistémique* – de satisfaire sa « curiosité » sous toutes ses formes. Si la « victime » est passive ou indifférente à un sujet x , alors il n'est pas nécessaire de préparer du matériau pour le sujet x . (Là où cela ne démange pas, pas la peine de gratter ?) Le monde nous fournit un déluge ininterrompu d'informations qui bombardent nos sens. Quand nous nous concentrons sur l'information que nous recevons ou qui est à notre disposition en continu, nous succombons souvent à l'illusion que tout doit être utilisé, tout le temps. Mais nos capacités d'utiliser l'information et nos appétits épistémiques sont limités. Si nos cerveaux ne peuvent satisfaire nos faims épistémiques que lorsqu'elles surgissent, nous n'aurons jamais l'occasion de nous plaindre. Nous ne serons

jamais capables de dire, en fait, que nos cerveaux nous fournissent moins que tout ce qui est à notre disposition dans le monde.

Pour le moment, on s'est contenté d'introduire ce principe économique, on ne l'a pas établi. Comme nous le verrons, le cerveau ne prend pas toujours cette option, mais il est important de ne pas manquer cette possibilité. On n'a pas assez reconnu combien ce principe peut permettre de résoudre de vieilles énigmes.

Aperçu d'ensemble

Dans les chapitres qui vont suivre, je tenterai d'expliquer la conscience. Plus précisément, j'expliquerai les divers phénomènes qui composent la conscience, en montrant comment ils sont tous des effets physiques des activités du cerveau, comment ces activités ont évolué et comment elles donnent lieu à des illusions quant à leurs pouvoirs et quant à leurs propriétés. Il est très difficile d'imaginer comment votre esprit pourrait être votre cerveau, mais ce n'est pas impossible. Pour l'imaginer, il faut connaître de nombreuses découvertes scientifiques concernant le fonctionnement du cerveau, mais plus important encore, il faut apprendre de nouvelles façons de penser. Apprendre des faits nouveaux nous aide à imaginer de nouvelles possibilités, mais les découvertes et les théories des neurosciences ne sont pas suffisantes – même les praticiens des neurosciences sont souvent perplexes face à la conscience. Dans le but d'étendre vos pouvoirs d'imagination, je fournirai, en même temps que les faits scientifiques, une série d'histoires, d'analogies, d'expériences de pensée et d'autres techniques destinées à vous donner de nouvelles perspectives, à briser de vieilles habitudes de pensée et à vous aider à organiser les faits en une vision d'ensemble cohérente, étonnamment différente de la conception traditionnelle de la conscience sur laquelle nous nous appuyons. L'expérience de pensée du cerveau dans la cuve et l'analogie avec le jeu de la psychanalyse sont des exercices d'entraînement en vue de notre tâche principale, qui consiste à esquisser une théorie des mécanismes biologiques et *une certaine façon de penser* à propos de ces mécanismes, qui vous permettra de voir comment on peut résoudre les paradoxes et les mystères traditionnels de la conscience.

Dans la première partie, je passe en revue les problèmes de

la conscience et je mets en place quelques méthodes. La tâche est plus difficile qu'elle n'en a l'air. Bien des problèmes rencontrés par les autres théories viennent de ce qu'elles ne partent pas du bon pied et qu'elles essaient de deviner trop tôt les réponses aux Grandes Questions. Les hypothèses qui se profilent derrière ma théorie jouent un rôle important dans ce qui suit. Elles me permettent de *différer* bien des énigmes philosophiques auxquelles se heurtent traditionnellement les théoriciens, jusqu'au moment où nous pourrons esquisser une théorie empiriquement fondée, qu'on trouvera dans la seconde partie.

Le modèle des Versions Multiples de la conscience ébauché dans la seconde partie vise à concurrencer le modèle traditionnel, que j'appelle Théâtre Cartésien. Il requiert une révision radicale de la notion courante de « flux de la conscience », et au départ, il choque beaucoup l'intuition, mais il s'imposera progressivement à vous, au fur et à mesure que vous comprendrez comment, au sujet du cerveau, il parvient à rendre compte de faits que les philosophes – et les savants – ont jusqu'à présent ignorés. L'examen détaillé de l'évolution de la conscience permet de mieux comprendre certaines propriétés de notre cerveau qui seraient troublantes si on ne les expliquait pas ainsi. La seconde partie fournit également une analyse du rôle du langage dans la conscience humaine et de la relation du modèle des Versions Multiples à certaines conceptions courantes de l'esprit, ainsi qu'à d'autres travaux théoriques appartenant au champ multidisciplinaire des sciences cognitives. Cela permet de résister au simplisme de la conception traditionnelle et d'essayer de trouver de nouveaux fondements.

Armés des outils qui guideront nos imaginations, dans la troisième partie, nous pourrons ainsi affronter les mystères traditionnels de la conscience : les propriétés étranges du « champ phénoménal », la nature de l'introspection, les qualités (ou *qualia*) des états de l'expérience, la nature du moi ou de l'*ego*, sa relation aux pensées et aux sensations et la conscience des créatures non humaines. On découvrira alors que les paradoxes qui encombre les débats philosophiques traditionnels sur ces sujets proviennent de notre *manque d'imagination*, pas de notre manque de « compréhension » de ces phénomènes. C'est ainsi que nous pourrons résoudre ces mystères.

Ce livre présente une théorie qui est à la fois empirique et philosophique. Puisqu'elle doit répondre à des conditions très

diverses, j'ai inclus deux appendices qui traitent brièvement de deux défis techniques qu'on rencontre à la fois du point de vue scientifique et du point de vue philosophique. Dans le chapitre suivant, nous abordons la question de savoir en quoi peut consister une explication de la conscience et si nous devons vraiment aspirer à résoudre les mystères de la conscience.

NOTES

1. Les dates entre parenthèses font référence à la liste des ouvrages qu'on trouve en bibliographie.

2. Le terme *explosion combinatoire* vient de l'informatique, mais le phénomène était connu bien avant l'existence des ordinateurs. Par exemple, dans la fable de l'empereur qui accepte de récompenser le paysan qui lui a sauvé la vie en mettant un grain de riz sur la première case de l'échiquier, deux grains sur la seconde, quatre sur la troisième, et ainsi de suite, en doublant le total pour chacune des soixante-quatre cases. À la fin, il doit à l'astucieux paysan des millions de milliards de grains de riz (2^{64} pour être exact). Plus près de notre exemple, on trouve la technique employée par les romanciers français de « l'aléatoire », qui écrivent des romans dans lesquels, une fois qu'il a lu le chapitre 1, le lecteur tire à pile ou face et lit le chapitre 2a ou le chapitre 2b, selon le résultat, puis lit ensuite le chapitre 3aa, 3ab, 3ba ou 3bb, et ainsi de suite, en tirant à pile ou face à la fin de chaque chapitre. Ces romanciers ont très vite réalisé qu'ils avaient intérêt à minimiser le nombre de choix s'ils voulaient éviter une explosion de fiction empêchant qui que ce soit de ramener le livre de chez le libraire à son domicile.

3. Le développement de systèmes de « Réalité Virtuelle » destinés à la récréation et à la recherche connaît en ce moment un boom. Les résultats sont impressionnants : des gants habillés électroniquement fournissent une interface permettant de « manipuler » des objets virtuels, et des montages visuels permettent d'explorer des environnements virtuels d'une complexité considérable. Ces systèmes ont néanmoins des limites et confirment ce que j'ai dit : ce n'est qu'en combinant des répliques physiques et par schématisation (une représentation relativement grossière) que l'on peut maintenir une illusion persistante. Et même dans le meilleur des cas, ces expériences font attendre une surréalité virtuelle et non des objets que l'on peut considérer durablement comme réels. Si vous voulez réellement faire croire à quelqu'un qu'il est dans une cage avec un gorille, il vaudra mieux, et pour longtemps encore, faire appel à un acteur déguisé en gorille.

4. En français dans le texte (NdT).

5. Pour une analyse plus détaillée des problèmes du libre arbitre, du contrôle de la lecture de ce qu'il y a dans le cerveau, voir mon livre *Elbow Room : The Varieties of Free Will Worth Wanting*, 1984, en particulier les chapitres 3 et 4.

6. Les tests empiriques suggèrent que le jeu a plus de chances de produire une bonne histoire si vous favorisez légèrement les réponses positives, en plaçant dans p / q la ligne de division alphabétique entre oui et non.

7. En français dans le texte (NdT).

8. Littéralement : ferme dans le Vermont et banque à Porto Rico (NdT).

9. Allusion à une maxime bien connue du philosophe américain W.V.O. Quine (NdT).

PREMIÈRE PARTIE

Problèmes et méthodes

Chapitre 1

Expliquer la conscience

La boîte de Pandore : la conscience doit-elle être démystifiée ?

Voici encore des arbres et je connais leur rugueux,
de l'eau et j'éprouve sa saveur. Ces parfums d'herbe
et d'étoiles, la nuit, certains soirs où le cœur se
détend, comment nierais-je ce monde dont j'éprouve
la puissance et les forces ? Pourtant, toute la science
de cette terre ne me donnera rien qui puisse m'as-
surer que ce monde est à moi. Vous me le décrivez
et vous m'apprenez à le classer. Vous énumérez ses
lois et dans ma soif de savoir je consens qu'elles
soient vraies. Vous démontez son mécanisme et mon
espoir s'accroît.... Qu'avais-je besoin de tant d'ef-
forts ? Les lignes douces de ces collines et la main
du soir sur ce cœur agité m'en apprennent bien
plus.

Albert Camus, *Le Mythe de Sisyphe*, 1942.

Doux est l'héritage que la Nature nous apporte,
Notre intellect occupé
Déforme les formes merveilleuses des choses
Nous tuons pour les disséquer.

William Wordsworth,
« The Tables Turned », 1798.

La conscience humaine est peut-être le dernier mystère qui
reste. Un mystère est un phénomène dont on ne sait quoi penser

— dont on ne sait pas encore ce qu'il faut penser. D'autres grands mystères ont existé : celui de l'origine de l'univers, celui de la vie et de la reproduction, celui du dessein qu'on peut lire dans la nature, celui du temps, de l'espace et de la gravitation. Ce n'étaient pas seulement des zones d'ignorance scientifique ; c'étaient des sujets de perplexité extrême et d'étonnement. Nous n'avons toujours pas de réponse ultime aux questions que posent la cosmologie et la physique des particules, la génétique moléculaire et la théorie de l'évolution, mais nous savons au moins les aborder. Les mystères ne se sont pas évanouis, mais on les a apprivoisés. Ils ont cessé de dépasser nos efforts de compréhension des phénomènes, parce que nous savons à présent distinguer les fausses questions des vraies, et même s'il se révèle que nous nous trompons du tout au tout dans certaines des réponses les plus couramment admises, nous savons où chercher pour en trouver de meilleures.

Mais avec la conscience, nous sommes toujours terriblement embourbés. La question de la nature de la conscience est aujourd'hui le seul sujet qui laisse souvent perplexes et muets les penseurs les plus subtils. Et, comme pour tous les autres mystères antérieurs, beaucoup soutiennent — et espèrent — qu'on ne pourra jamais démystifier la conscience.

Après tout, les mystères sont excitants, et ils contribuent à rendre la vie amusante. Personne n'apprécie le gâcheur qui donne la clef de l'énigme à ceux qui font la queue pour aller voir un film. À partir du moment où on a révélé le pot aux roses, on ne peut plus retrouver l'état délicieux de mystification qui nous avait d'abord envoûtés. Par conséquent, le lecteur doit être sur ses gardes. Si je réussis dans ma tentative pour expliquer la conscience, ceux qui iront au bout de leur lecture échangeront le mystère contre des rudiments de connaissance scientifique concernant la conscience. Cet échange ne sera pas honnête, au goût de certains. Comme certaines personnes considèrent la démystification comme une forme de profanation, je m'attends à ce qu'ils considèrent d'emblée ce livre comme un acte de vandalisme intellectuel, comme un assaut mené contre le dernier sanctuaire de l'espèce humaine. Je voudrais leur faire changer d'avis.

Camus suggère qu'il n'a pas besoin de la science, car il peut en apprendre plus des douceurs des collines et de la main du soir. Je n'ai rien à lui objecter, vues les questions que Camus se

pose. La science ne répond pas à toutes les bonnes questions. Pas plus que la philosophie. Mais c'est précisément pour cette raison que les phénomènes de la conscience, qui sont problématiques en eux-mêmes indépendamment des questions que se pose Camus, n'ont pas besoin d'être protégés de la science ou du type d'investigation philosophique démystifiante dans laquelle nous nous embarquons. Il arrive que des gens, qui ont peur que la science « tue pour disséquer », comme disait Wordsworth, soient attirés par des doctrines philosophiques qui offrent une forme de garantie contre ce genre d'invasion. Les appréhensions qui les motivent sont fondées, quelles que soient les forces et les faiblesses des doctrines ; il se *pourrait* bien, en fait, que la démystification de la conscience soit une grande perte. Je soutiendrai seulement que cela *ne se produira pas* : les pertes, s'il y en a, seront compensées par les gains en compréhension – à la fois scientifiques et sociaux, à la fois théoriques et moraux – qu'une bonne théorie de la conscience peut nous fournir.

Comment, pourtant, *pourrions-nous* regretter la démystification de la conscience ? Cela pourrait être comme la perte de l'innocence enfantine, qui est définitive, même si elle est bien dédommagée. Pensons à ce qui arrive à l'amour, par exemple, quand nous devenons des êtres plus sophistiqués. Nous comprenons comment un chevalier à l'époque de la chevalerie pouvait vouloir sacrifier sa vie pour l'honneur d'une princesse à qui il avait seulement parlé. C'est une idée qui me transportait quand j'avais onze ou douze ans. Mais ce n'est pas le genre d'état d'esprit dans lequel un adulte d'aujourd'hui pourrait aisément se trouver. Il y eut un temps où les gens parlaient d'amour et l'envisageaient de façons qui sont pour nous pratiquement impossibles à concevoir – sauf pour les enfants et ceux qui parviennent à mettre de côté leurs connaissances d'adultes. Nous aimons tous dire à ceux que nous aimons que nous les aimons et entendre de leur bouche que nous sommes aimés. Mais, dès que nous sommes devenus adultes, nous ne sommes plus tout à fait aussi sûrs qu'avant, quand nous étions enfants et que l'amour était une chose simple, de savoir ce que cela veut dire.

Est-ce que nous gagnons ou est-ce que nous perdons à changer ainsi de perspective ? Le changement n'est pas uniforme, bien entendu. Alors que les adultes naïfs continuent de porter des romans gothiques au sommet de la liste des best-sellers, nous autres, lecteurs sophistiqués, considérons que nous sommes deve-

nus tout à fait immunisés contre les effets visés par ce genre de livres : ils nous font rigoler, ils ne nous font pas pleurer. Ou bien s'ils nous font pleurer — comme cela arrive, malgré nous — nous sommes embarrassés de découvrir que nous sommes toujours capables de nous laisser avoir par des trucs aussi faciles. Nous ne pouvons pas aisément partager les sentiments de l'héroïne qui passe son temps à s'inquiéter de savoir si elle a trouvé son « véritable amour », comme si c'était une sorte de substance particulière (de l'or émotionnel, par opposition à du laiton ou à du cuivre émotionnel). Ce mûrissement n'est pas seulement le fait de l'individu. Notre culture est devenue plus sophistiquée. Tout du moins la sophistication, quelle que soit sa valeur, est-elle plus largement répandue à travers la culture. Il en résulte que nos concepts de l'amour ont changé, et que ces modifications s'accompagnent de changements de sensibilité qui empêchent à présent d'éprouver certaines des expériences qui ébranlaient, dévastaient ou galvanisaient nos ancêtres.

Il arrive quelque chose de semblable à la conscience. Nous parlons tous, aujourd'hui, de nos décisions conscientes et de nos habitudes inconscientes, des expériences conscientes que nous éprouvons (par opposition, par exemple, aux distributeurs automatiques de billets, qui n'ont pas d'expériences de ce genre). Mais nous ne sommes plus tout à fait aussi sûrs de savoir ce que nous voulons dire quand nous disons ce genre de choses. Alors qu'il existe toujours des penseurs qui font semblant de croire que la conscience est une sorte de chose pure et précieuse (comme l'amour, comme l'or), une chose qui est simplement « évidente » et très, très spéciale, on soupçonne de plus en plus que ce n'est là qu'une illusion. Peut-être les divers phénomènes qui conspirent à créer le sentiment qu'il existe un phénomène unique et mystérieux n'ont-ils pas plus d'unité ultime ou essentielle que les divers phénomènes qui contribuent à nous donner le sentiment que l'amour est une chose simple.

Comparons l'amour et la conscience avec deux phénomènes assez différents, les maladies et les tremblements de terre. Nos concepts des maladies et des tremblements de terre ont eux aussi beaucoup changé durant les derniers siècles. Mais les maladies et les tremblements de terre sont des phénomènes très largement indépendants des concepts que nous en avons. Le fait de changer d'avis au sujet des maladies n'a pas par soi-même fait disparaître les maladies ou ne les a pas rendues moins fréquentes, bien que

cela ait produit des changements dans la pratique de la médecine et dans la santé publique qui ont radicalement changé les formes de manifestation des maladies. De même, il se peut que les humains réussissent un jour à contrôler plus ou moins les tremblements de terre ou tout au moins parviennent à les prédire. Mais dans l'ensemble, l'existence des tremblements de terre n'est pas affectée par les attitudes que nous adoptons à leur sujet ou par les concepts que nous en avons. Avec l'amour, c'est différent. Il n'est plus possible aux gens sophistiqués de « tomber amoureux » à la manière dont c'était possible autrefois – tout simplement parce qu'ils ne peuvent pas croire à *ces* façons de tomber amoureux. Il ne m'est plus possible, par exemple, – à moins de « retomber en adolescence » – d'avoir un pur coup de foudre d'adolescent et au cours de ce processus, d'oublier ou d'abandonner une bonne partie de ce que je crois savoir. Heureusement, je peux croire encore en d'autres formes d'amour. Mais que se passerait-il s'il n'y en avait pas ? Pour simplifier à l'excès pour le moment, je dirais que l'amour est l'un de ces phénomènes qui *dépendent de leurs concepts*. Il en existe d'autres ; l'argent en est un exemple clair. Si tout le monde oubliait ce qu'est l'argent, il n'y aurait plus d'argent du tout : il n'y aurait que des tas de bouts de papiers avec des gravures dessus, de disques de métal frappés, des relevés informatiques de comptes en banque, des banques de granit et de marbre, mais pas d'argent, pas d'inflation ou de déflation, pas de taux d'intérêts et même pas de *valeur monétaire*. La propriété même, qui explique – sans que rien d'autre ne puisse l'expliquer – la trajectoire de ces différents morceaux de papier imprimé d'une main à une autre au cours de différents actes et échanges, s'évaporerait.

Selon la conception de la conscience que je développerai dans ce livre, la conscience se révélera être, comme l'amour et l'argent, un phénomène qui dépend en fait dans des proportions surprenantes des concepts qui lui sont associés. Bien que, comme l'amour, elle ait un fondement biologique bien précis, certaines de ses caractéristiques sont, comme l'argent, des créations culturelles, pas simplement des traits inhérents à la structure physique de ses réalisations. Par conséquent, si ma thèse est correcte, et si je réussis à rejeter certains de ces concepts, je menacerai d'extinction tous les phénomènes de conscience qui en dépendent. Allons-nous entrer dans la période post-consciente de la concep-

tualisation humaine ? N'est-ce pas quelque chose qu'il nous faut craindre ? Est-ce même concevable ?

Si le concept de conscience devait « tomber aux mains de la science » qu'advierait-il de notre sens de l'action morale et de notre libre arbitre ? Si l'expérience consciente devait d'une manière ou d'une autre se « réduire » à de la simple matière en mouvement, qu'advierait-il de notre expérience de l'amour, de la douleur, des rêves et de la joie ? Si les êtres humains conscients étaient « seulement » des objets matériels inanimés, comment ce que nous leur faisons pourrait-il être juste ou injuste ? Voilà quelques-unes des peurs qui nourrissent la résistance et qui distraient la concentration de ceux qu'on met en présence de tentatives pour expliquer la conscience.

Je suis sûr que ces craintes sont infondées, mais ce n'est pas évident. Elles sont l'enjeu de la confrontation de théories et d'arguments qui va commencer. Certains arguments puissants, tout à fait indépendants de ces craintes, s'opposent au type de théorie scientifique matérialiste que je m'appête à proposer. Je reconnais qu'il m'incombe de démontrer non seulement que ces arguments sont erronés, mais aussi que si l'on accepte largement ma conception de la conscience, cela n'aurait en aucun cas ces conséquences désastreuses. (Si j'avais découvert que cette conception avait des chances de produire ces effets, qu'aurais-je fait ? Je n'aurais pas écrit *ce* livre. Mais, à part cela, je ne sais pas.)

Si nous voyons les choses du bon côté, rappelons-nous ce qui s'est passé au moment des démystifications précédentes. Notre étonnement n'a pas pour autant diminué ; au contraire, nous découvrons des beautés encore plus profondes et des visions encore plus étonnantes de la complexité de l'univers que celles que les protecteurs du mystère ont pu jamais concevoir. La « magie » des visions antérieures servait seulement à couvrir de purs manques d'imagination, une esquivе ennuyeuse enrichie par le concept de *deus ex machina*. De fières divinités conduisant des chars dorés à travers les cieux apparaissent comme de simples personnages de bande dessinée, si on les compare à la merveilleuse étrangeté de la cosmologie contemporaine, et les complexités sans cesse croissantes de la machinerie de l'ADN rendent l'*élan vital* à peu près aussi intéressant que la terrible kryptonite de Superman. Quand nous comprendrons la conscience, quand le mystère s'éva-

nouira, la conscience sera différente. Mais la beauté sera toujours là. Et il restera plus de place que jamais pour l'émerveillement.

Le mystère de la conscience

En quoi consiste donc le mystère ? Que peut-il y avoir de plus évident ou de plus certain pour chacun d'entre nous que le fait qu'il est un sujet d'expérience consciente, qui a des perceptions et des sensations, qu'il éprouve des douleurs, qu'il a des idées et qu'il délibère consciemment ? Cela semble indéniable. Mais qu'est-ce que la conscience elle-même, si c'est une chose qui existe dans le monde ? Comment des êtres vivants physiques appartenant au monde physique peuvent-ils produire de tels phénomènes ? C'est là le mystère.

Le mystère de la conscience peut se présenter de multiples façons. Il m'a frappé à nouveau avec une force particulière il y a quelques jours alors que j'étais assis dans un fauteuil à bascule en train de lire un livre. Je venais apparemment de détacher mes yeux du livre, et j'avais d'abord tourné machinalement mon regard en direction de la fenêtre, perdu dans mes pensées, quand la beauté de ce qui m'entourait me détourna de mes préoccupations théoriques. Un rayon de soleil mordoré pénétrait par la fenêtre en ce début de printemps ; dans le jardin, les milliers de branches et de ramures de l'érable étaient encore bien visibles à travers une brume de bourgeons verts formant une élégante trame merveilleuse de complexité. La vitre est en verre ancien, et il y a une ride qu'on décèle à peine. Au fur et à mesure de mon balancement sur la chaise, cette imperfection du verre produisit une vague de tressautements synchronisés qui allaient et venaient à travers le delta de branches. Ce mouvement régulier se superposait avec une vivacité remarquable au chatoiement plus chaotique des ramures et des branchages qu'ébranlait la brise.

C'est alors que je remarquai que ce métronome visuel des branches de l'arbre suivait étroitement le rythme du concerto grosso de Vivaldi que j'étais en train d'écouter en « musique d'arrière-fond » pendant ma lecture. D'abord, je pensai qu'il était évident que j'avais dû synchroniser inconsciemment mon balancement sur la chaise avec la musique — tout comme on bat le rythme inconsciemment du pied. Mais le balancement des fau-

teuils à bascule suit des fréquences plutôt limitées. Le synchronisme était sans doute avant tout une pure coïncidence, légèrement influencée par ma préférence inconsciente pour un rythme net, me permettant de garder la mesure.

Je me mis à songer un instant à d'hypothétiques processus cérébraux qui pourraient expliquer la façon dont nous ajustons inconsciemment notre comportement, y compris le comportement de nos yeux et de nos facultés d'attention, pour « synchroniser » la « bande-son » avec l'« image ». Mais ces songeries furent à leur tour interrompues quand je réalisai soudain quelque chose. *Ce que j'étais en train de faire* – la combinaison d'expérience vécue et de pensée que je viens de décrire à partir de mon point de vue privilégié à la première personne, était bien plus difficile à « modéliser » que les processus inconscients, en coulisses, qui se déroulaient indubitablement *en moi* et qui formaient en quelque sorte les conditions causales de ce que j'étais en train de faire. La machinerie opérant en coulisses était relativement facile à comprendre ; le plus troublant était ce qui se passait au premier plan, les événements placés devant les feux de la rampe. Ma pensée consciente, en particulier le plaisir que je prenais au mélange de rayons de soleil, de violons vivaldiens ensoleillés, d'ondulations des branches, le plaisir que je prenais à penser à tout cela, comment *tout cela* pouvait-il être seulement quelque chose de physique qui se produisait dans mon cerveau ? Comment une combinaison d'événements électrochimiques ayant lieu dans mon cerveau pouvait-il en quelque sorte s'ajouter à la façon délicieuse dont des centaines de bourgeons se balançaient en mesure avec la musique ? Comment un processus de traitement de l'information localisé dans mon cerveau pouvait-il *être* la délicate douceur de la lumière du soleil que je sentais tomber sur moi ? En l'occurrence, comment un événement dans mon cerveau pouvait-il *être* l'image mentale fugitive que j'avais... d'un autre événement de traitement de l'information se déroulant dans mon cerveau ? Cela semble vraiment impossible.

Tout se passe comme si les événements qui *sont* mes pensées conscientes ne pouvaient être des événements du cerveau, mais devaient être *quelque chose d'autre*, qui a été causé ou produit par des événements et qui, sans aucun doute, s'y ajoute, est fait d'une substance différente et est situé dans un espace différent. Et pourquoi pas ?

Les attraites de la substance mentale

Voyons ce qui se passe quand on emprunte cette voie, qui est indéniablement tentante. Je veux tout d'abord faire une expérience simple. Elle consiste à vous demander de fermer les yeux, d'imaginer quelque chose. Puis, une fois que vous avez formé votre image mentale et l'avez vérifiée attentivement, répondez aux questions qui suivent. Ne lisez pas les questions tant que vous n'avez pas suivi l'instruction : quand vous fermez les yeux, imaginez, avec autant de détails que possible, une vache violette. C'est fait ? Alors :

1. Votre vache était-elle tournée vers la gauche, vers la droite, ou était-elle de face ?
2. Était-elle en train de ruminer ?
3. Son pis était-il visible ?
4. Était-elle d'un violet relativement pâle, ou d'un violet profond ?

Si vous avez suivi les instructions, vous avez probablement pu répondre aux quatre questions sans avoir à consulter rétrospectivement votre image. Si vous avez trouvé que les quatre questions étaient terriblement exigeantes, alors c'est que vous ne vous êtes même pas soucié d'imaginer une vache violette. Vous avez juste pensé paresseusement « je suis en train d'imaginer une vache » ou « appelons *cette* chose que j'imagine une vache violette », ou bien vous avez fait quelque chose de vague de ce genre.

Passons maintenant à un second exercice : fermez les yeux et imaginez, avec autant de détails que possible, une *vache jaune*.

Cette fois-ci vous pouvez probablement répondre sans problème aux trois premières questions. Vous pourrez dire avec certitude quelle sorte de jaune — pastel ou couleur de beurre ou ocre — couvrirait les flancs de la vache que vous avez imaginée. Mais cette fois-ci, je voudrais considérer une question différente :

5. Quelle est la différence entre imaginer une vache violette et imaginer une vache jaune ?

La réponse va de soi : la première vache imaginée est violette et la seconde est jaune. Il pourrait y avoir d'autres différences, mais celle-là est la principale. Le problème est que puisque ces vaches ne sont que des vaches imaginées, et non pas des vaches réelles, plutôt que des vaches réelles, ou des images de vaches

peintes sur une toile, ou des formes de vaches sur un écran de télévision en couleur, il est difficile de dire ce qui pourrait être violet dans le premier cas et jaune dans le second. Il n'y a rien dans votre cerveau (ou dans votre globe oculaire) qui ait la forme approximativement d'une vache et qui se révèle violet dans un cas et jaune dans l'autre. Même si c'était le cas, ce ne serait pas très utile, parce qu'il fait noir comme dans un four au fond de votre boîte crânienne, et qui plus est, vous n'avez pas d'yeux à l'intérieur pour voir des couleurs.

Il existe dans votre cerveau des événements qui sont étroitement associés à vos actes particuliers d'imagination. Il n'est donc pas hors de question que dans un futur proche, un neurophysiologiste, analysant les processus qui se sont déroulés dans votre cerveau en réponse à mes instructions, soit capable de les déchiffrer au point d'être en mesure de confirmer ou d'infirmer vos réponses aux questions 1 à 4.

« La vache était-elle tournée vers la gauche ? Nous le pensons. La trame d'excitation neuronale correspondant à la tête de la vache s'accordait avec le quadrant visuel supérieur gauche, et nous avons observé des signaux de détection de mouvement oscillatoire d'un Hertz qui ont suggéré la rumination, mais nous n'avons pu détecter d'activité dans les groupes de représentations du complexe-pis, et, après avoir calibré les potentiels suscités avec les profils de détection de couleur du sujet, nous formons l'hypothèse que le sujet ment au sujet de la couleur : la vache imaginée était presque certainement brune. »

Supposez que tout cela soit vrai. Supposez qu'on en soit arrivé à pouvoir lire dans l'esprit des gens. Même alors, semble-t-il, le mystère demeure : qu'est-ce qui est brun quand vous imaginez une vache brune ? Pas l'événement dans le cerveau que les savants ont calibré avec votre activité-de-faire-l'expérience-du-brun. Le type de neurones impliqués et leur localisation, leurs connexions avec d'autres parties du cerveau, la fréquence ou amplitude d'activité, les neurotransmetteurs chimiques émis – aucune de ces propriétés n'est la propriété même de la vache « dans votre imagination ». Et puisque vous avez bien imaginé une vache (vous n'êtes pas en train de mentir – les savants peuvent même vous le confirmer), une vache imaginée est venue à l'existence à ce moment précis. La vache imaginée doit se trouver non pas dans le milieu de la substance cérébrale, mais dans le milieu de la.... substance mentale. De quoi d'autre pourrait-il s'agir ?

La substance mentale, donc, doit être « ce dont les rêves sont faits », et elle a apparemment quelques propriétés remarquables. Nous avons relevé l'une d'elles au passage, mais elle se laisse très difficilement définir. Pour commencer, disons que la substance mentale a toujours *un témoin*. Le problème avec les événements cérébraux, comme nous l'avons remarqué, c'est que quel que soit le degré de leur « correspondance » avec les événements qui interviennent durant le cours de notre conscience, ils ont fatalement un inconvénient : *il n'y a personne là-dedans pour les regarder*. Il n'y a d'habitude personne pour témoigner des événements qui apparaissent dans votre cerveau, tout comme pour les événements qui se produisent dans votre estomac ou dans votre foie ; le fait qu'il y ait quelqu'un ou pas pour en témoigner ne change rien au fait qu'ils aient lieu. En revanche les événements qui se produisent dans la conscience sont « par définition » des événements qui ont un témoin : ils font l'objet d'*expériences* pour un *sujet qui a des expériences*. Le fait qu'ils fassent ainsi l'objet d'expériences est ce qui fait d'eux ce qu'ils sont : *des événements conscients*. Un événement dont on fait l'expérience ne peut être suspendu tout seul en l'air : il doit être un objet d'expérience *pour quelqu'un*. Pour qu'il se produise une pensée, quelqu'un (un esprit) doit la penser ; pour qu'une douleur se produise, quelqu'un doit l'éprouver ; et pour qu'une vache violette vienne à l'existence « dans notre imagination », il doit y avoir quelqu'un pour l'imaginer.

Le problème qui se pose avec les cerveaux, semble-t-il, c'est que quand vous les regardez, vous découvrez qu'*il n'y a personne à la maison*. Aucune partie du cerveau n'est le penseur qui pense ou le sujet qui éprouve la sensation, et le cerveau tout entier ne semble pas être un meilleur candidat pour jouer ce rôle particulier. C'est un sujet qui nous glisse entre les doigts. Les cerveaux pensent-ils ? Les yeux voient-ils ? Ou les gens voient-ils avec leurs yeux et pensent-ils avec leurs cerveaux ? Y a-t-il une différence ? Est-ce seulement un point trivial de « grammaire » ou cela révèle-t-il une source de confusion ? L'idée selon laquelle un *moi* (ou une personne, ou en l'occurrence une âme) est distinct d'un cerveau est profondément ancrée dans nos façons de parler, et par conséquent dans nos manières de penser.

J'ai un cerveau.

C'est là quelque chose qui semble être parfaitement incontestable. Et cela ne semble pas vouloir dire seulement que :

Ce corps a un cerveau (et un cœur, et deux poumons, etc.)
ou

Ce cerveau se possède lui-même.

Il est tout à fait naturel de penser que « le moi et son cerveau » (Popper et Eccles, 1977) sont deux choses distinctes, avec des propriétés différentes, quel que soit le degré de leur dépendance mutuelle. Si le moi est distinct du cerveau, il semble qu'il doive être fait de substance mentale. En latin, un être pensant est une *res cogitans*, expression rendue fameuse par Descartes, qui a donné ce qu'il considérerait comme une preuve inébranlable du fait que lui-même, étant manifestement une chose pensante, ne pouvait être son cerveau. Voici l'une des versions de cette preuve. Elle est certainement convaincante :

« Puis, examinant avec attention ce que j'étais, et voyant que je ne pouvais feindre que je n'avais aucun corps et qu'il n'y avait aucun monde ni aucun lieu où je fusse, mais que je ne pouvais pas feindre pour cela que je n'étais point, et qu'au contraire, de cela même que je pensais à douter de la vérité des autres choses, il suivait très évidemment et très certainement que j'étais ; au lieu que, si j'eusse seulement cessé de penser, encore que tout le reste de ce que j'avais imaginé eût été vrai, je n'avais aucune raison de croire que j'eusse été, je connus de là que j'étais une substance dont toute l'essence ou la nature n'est que de penser, et qui, pour être, n'a besoin d'aucun lieu ni ne dépend d'aucune chose matérielle. » (*Discours de la méthode*, 1637.)

Nous avons donc découvert deux sortes de choses que l'on pourrait vouloir construire avec de la substance mentale : la vache violette qui n'est pas dans le cerveau, et la chose qui produit la pensée. Mais nous pouvons vouloir attribuer encore d'autres choses à la substance mentale.

Supposons que les propriétaires d'un vignoble aient décidé de remplacer leurs taste-vin par une machine. Un « système expert » fonctionnant sur ordinateur et permettant de contrôler la qualité des vins et de les classer est presque à la portée de la technologie existante. Nous en savons assez sur la chimie du vin pour fabriquer les transducteurs qui pourraient remplacer les papilles et les récepteurs olfactifs de l'épithélium (en fournissant le « matériau brut » – les stimuli d'entrée – pour le goût et l'odeur). On ne sait pas

précisément comment ces entrées se combinent et interagissent pour produire nos expériences, mais on fait des progrès dans cette direction. On a fait des avancées bien plus importantes dans le domaine de la vision. Les recherches sur la vision des couleurs suggèrent qu'imiter l'idiosyncrasie humaine, la délicatesse et la fiabilité du composant de la machine chargé de jouer les couleurs serait un défi technique considérable, mais ce n'est pas hors de question. Nous pouvons donc d'emblée imaginer que nous pourrions utiliser les sorties avancées de ces transducteurs sensoriels et leur système de comparaison pour approvisionner des routines de classification en descriptions et en évaluations élaborées. Versez un échantillon de vin dans l'entonnoir, et après quelques minutes ou quelques heures, la machine imprimera un rapport chimique, accompagné du commentaire suivant : « Pinot flamboyant et pourpre ; mais peu de corps » ou quelque chose de ce genre. Ce genre de machine pourrait même être capable de meilleures performances que certains taste-vin humains selon tous les tests raisonnablement exacts et cohérents que pourraient imaginer des fabricants humains de vin. Mais il ne fait pas de doute qu'aussi « sensibles » et « précis » que puisse devenir un tel système, il semble qu'il n'aura jamais et ne pourra jamais éprouver la même chose que ce qui se passe quand *nous* goûtons un vin.

Est-ce si évident ? Selon les différentes idéologies regroupées sous le terme de *fonctionnalisme*, si vous reproduisiez la « structure fonctionnelle » tout entière du système cognitif du taste-vin humain (y compris sa mémoire, ses buts, ses aversions intimes, etc.), vous reproduiriez par là même *toutes* les propriétés mentales, y compris le plaisir, l'agrément et la saveur qui fait que beaucoup d'entre nous *apprécient* de boire du vin. En principe, cela ne fait pas de différence, nous dit le fonctionnaliste, si un système est fait de molécules organiques de silicone, tant qu'il *fait la même chose*. Les cœurs artificiels n'ont pas besoin d'être faits de tissus organiques, pas plus que les cerveaux artificiels – au moins en principe. Si toutes les fonctions de contrôle du cerveau du taste-vin humain peuvent être reproduites par des morceaux de silicone, notre plaisir à boire du vin se trouvera reproduit lui aussi.

Il se peut qu'une version ou une autre de fonctionnalisme triomphe en dernière instance (en fait, ce livre défendra une version du fonctionnalisme), mais cela semble inacceptable au premier abord. Il semble qu'une simple machine, quel que soit

le degré de précision avec lequel elle peut reproduire les processus cérébraux du taste-vin humain, est incapable d'apprécier un vin, une sonate de Beethoveen ou une partie de basket. Pour apprécier quelque chose, vous devez avoir une conscience, ce qu'une simple machine n'a pas. Mais il va de soi que le cerveau *est* une sorte de machine, un organe comme le cœur, les poumons ou les reins dont les pouvoirs s'expliquent en dernière instance de façon mécanique. Cela peut nous inviter à dire que ce n'est pas le cerveau qui fait apprécier : ce pouvoir-là est du ressort (ou est le privilège) de l'esprit. La reproduction de la machinerie du cerveau dans une machine faite de silicone ne produirait donc pas une appréciation réelle, mais au mieux une illusion ou un simulacre d'appréciation.

Il s'ensuit que l'esprit conscient n'est pas simplement le lieu où se trouvent les couleurs et les odeurs observées, et pas seulement la chose pensante. C'est le lieu où se produit l'appréciation. C'est lui qui arbitre en dernier ressort le fait que quelque chose peut avoir de l'importance pour nous. Peut-être est-ce là aussi une conséquence du fait que l'esprit conscient est supposé être la source de nos actions intentionnelles. Il est raisonnable de dire – n'est-ce pas ? – que si *faire des choses qui ont de l'importance* dépend de la conscience, le fait d'accorder une importance à quelque chose (éprouver du plaisir, apprécier, souffrir, se soucier au sujet de quelque chose) devrait dépendre aussi de la conscience. Si un somnambule blesse quelqu'un « inconsciemment », il n'est pas responsable parce qu'en un sens important ce n'est pas *lui* qui l'a fait ; ses mouvements corporels sont intimement impliqués dans les chaînes causales qui conduisent à la blessure, mais ils n'ont pas constitué une *action* qu'il aurait accomplie, pas plus que s'il avait blessé la personne en tombant de son lit. La simple complicité corporelle ne suffit pas à produire une action intentionnelle, pas plus que la complicité corporelle *sous le contrôle de structures du cerveau*, parce que le corps du somnambule est manifestement sous le contrôle de structures appartenant au cerveau du somnambule. Ce qui doit être ajouté en plus, c'est la conscience, cet ingrédient particulier qui transforme les simples *événements* en *choses que l'on fait* ¹.

Ce n'est pas la faute du Vésuve si son éruption tue votre bien-aimée. Pour pouvoir éprouver du ressentiment (Strawson, 1962) ou du dégoût à son encontre, vous devez vous convaincre que le Vésuve, contrairement aux opinions de vos contemporains, est

un agent conscient. Bien sûr, phénomène étrange, il est réconfortant, quand nous éprouvons de telles douleurs, de nous laisser aller à de telles attitudes, de protester contre la « furie » d'un ouragan, la malédiction du cancer qui frappe si injustement un enfant, ou de maudire « les dieux ». À l'origine, dire que quelque chose était « animé » par opposition à « inanimé » voulait dire que ce quelque chose avait une âme (*anima* en latin). Il peut être plus que simplement réconfortant de penser que les choses qui nous affectent fortement sont animées ; cela peut être une technique que nous devons à la manière dont nous sommes biologiquement conçus, un raccourci permettant à nos cerveaux soumis à la pression du temps d'organiser et de penser les choses auxquelles nous devons penser si nous devons survivre.

Il se peut que nous ayons une tendance innée à traiter spontanément toute chose qui change comme si elle avait une âme (Stafford, 1983 ; Humphrey, 1983b, 1986). Mais aussi naturelle que puisse être cette attitude, nous savons à présent qu'attribuer une âme (consciente) au Vésuve va trop loin. La question de savoir où tracer la limite est épineuse, nous y reviendrons. Mais pour nous, semble-t-il, la conscience est précisément ce qui nous distingue de simples « automates ». Les simples « réflexes corporels » sont « automatiques » et mécaniques ; ils peuvent impliquer des circuits dans le cerveau, mais ils ne requièrent pas l'intervention d'un esprit conscient. Il est très naturel de penser que nos propres corps sont de simples marionnettes que « nous » contrôlons « de l'intérieur ». Je fais saluer le public à ma marionnette en agitant mon doigt ; j'agite mon doigt en... en agitant quoi ? Mon âme ? Cette idée pose des problèmes bien connus, mais cela ne l'empêche pas d'avoir l'air vraie en un sens : à moins qu'il n'y ait un esprit conscient derrière mon acte, aucun agent *véritable* ne l'accomplit. Quand nous considérons notre esprit de cette façon, nous croyons découvrir le « moi interne », le « moi réel ». Ce moi réel n'est pas mon esprit ; c'est ce qui *possède* mon cerveau (« le moi et son cerveau »). Sur le bureau d'Harry Truman dans le Bureau Oval de la Maison Blanche il y avait un panneau célèbre : « Plus personne sur qui rejeter la responsabilité ². » Aucune partie du cerveau, semble-t-il, ne pourrait être là où *se trouvent les commandes*, ne pourrait être la source ultime de responsabilité morale située au début d'une chaîne de commandement.

Pour nous résumer, nous avons trouvé quatre raisons de croire qu'il existe une substance mentale. L'esprit conscient,

semble-t-il, ne peut pas simplement *être* le cerveau, ou une partie de celui-ci, parce que rien dans le cerveau ne pourrait :

1. être le milieu dans lequel se trouve la vache violette ;
2. être la chose pensante, le *Moi* dans « Je pense donc je suis » ;
3. apprécier le vin, haïr le racisme, aimer quelqu'un, être une source permettant de juger de l'importance d'une chose ;
4. agir de façon moralement responsable.

Une théorie acceptable de la conscience humaine doit rendre compte de ces quatre raisons de croire qu'il doit exister une substance mentale.

Pourquoi le dualisme est caduc

L'idée que l'esprit est différent du cerveau, qu'il est composé non de matière ordinaire mais d'une sorte particulière de substance s'appelle *dualisme*. Elle a, à juste titre, mauvaise réputation aujourd'hui, en dépit des conceptions influentes qui viennent d'être évoquées. Depuis la critique classique que Gilbert Ryle (1949) a développée de ce qu'il appelait le « dogme cartésien du fantôme dans la machine », les dualistes ont été sur la défensive³. La sagesse dominante, qui s'est exprimée sous diverses formes, est le *matérialisme* : il n'y a qu'une seule sorte de substance dont les choses sont faites – à savoir la matière, la substance physique de la physique, de la chimie et de la physiologie – et l'esprit n'est en un sens rien d'autre qu'un phénomène physique. En bref, l'esprit est le cerveau. Selon les matérialistes, nous pouvons (en principe !) expliquer tous les phénomènes mentaux en utilisant les mêmes principes physiques, les mêmes lois et les mêmes matériaux bruts que ceux qui sont suffisants pour expliquer la radioactivité, la dérive des continents, la photosynthèse, la reproduction, la nutrition et la croissance. Ce livre a précisément pour objectif principal d'expliquer la nature de la conscience sans jamais céder au chant des sirènes dualistes. Qu'est-ce qui, par conséquent, cloche dans le dualisme ? Pourquoi a-t-il si mauvaise réputation ?

L'objection habituelle que l'on adresse au dualisme était bien connue de Descartes lui-même au XVII^e siècle, et il est juste de dire que ni lui ni aucun dualiste par la suite n'y a jamais répondu de manière convaincante. Si l'esprit et le corps sont des substances distinctes, elles doivent cependant interagir ; les organes

des sens, à travers le cerveau, doivent *informer* l'esprit, ils doivent lui envoyer ou lui présenter des perceptions ou des idées d'un certain type ; et l'esprit doit ensuite, après avoir pensé ces choses, *diriger* le corps vers une action appropriée (y compris par le langage). C'est pourquoi cette théorie est parfois appelée interactionnisme cartésien ou dualisme interactionniste. Selon la formulation de Descartes, le lieu d'interaction avec le cerveau était la glande pinéale, ou l'*hypophyse*. Dans le diagramme que Descartes donnait lui-même, c'est le point ovale élargi qui apparaît au milieu de la tête.

Nous pouvons expliciter le problème posé par l'interactionnisme en ajoutant au diagramme de Descartes un schéma indiquant le reste de sa théorie.

La perception consciente de la flèche n'intervient qu'après que le cerveau a d'une certaine façon transmis son message à l'esprit, et le doigt du sujet ne peut pointer en direction de la flèche qu'après que l'esprit a commandé au corps. Comment, précisément, l'information se transmet-elle de la glande pinéale à l'esprit ? Comme nous n'avons pas la moindre idée (pour le moment) de ce que peuvent être les propriétés de la substance mentale, nous ne pouvons même pas envisager (pour le moment) comment

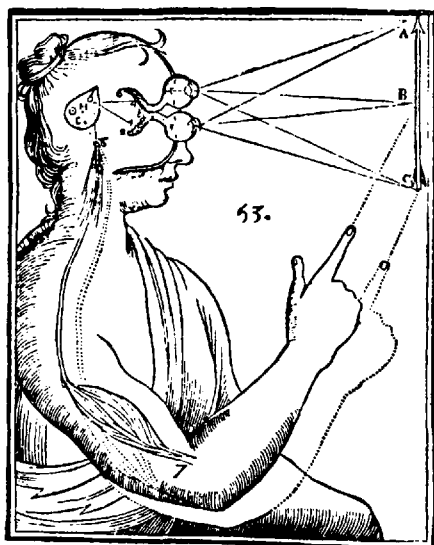


Figure 1.1

elle pourrait être affectée par des processus physiques qui émaneraient en quelque sorte du cerveau. Ignorons donc pour le moment les signaux dirigés vers le haut, et concentrons-nous sur ceux qui reviennent de l'esprit au cerveau. Ceux-ci, par hypothèse, ne sont pas physiques ; ce ne sont pas des ondes lumineuses ou sonores ni des rayons cosmiques ou des flux de particules subatomiques. Aucune énergie ou masse physique ne leur est associée. Comment, par conséquent, peuvent-ils faire une différence relativement à ce qui se passe dans les cellules du cerveau qu'ils doivent affecter, si l'esprit est supposé avoir une influence quelconque sur le corps ? Selon un principe fondamental de la physique, tout changement dans la trajectoire d'une entité physique est une accélération demandant une dépense d'énergie. D'où vient cette énergie ? C'est ce principe de conservation de l'énergie qui explique l'impossibilité des « machines à mouvement perpétuel », et c'est apparemment ce principe que viole le dualisme. On a discuté sans fin ce conflit entre la physique usuelle et le dualisme depuis l'époque de Descartes, et on considère en général que c'est le défaut inévitable et fatal du dualisme.

Comme on pourrait s'y attendre, on a exploré et exposé des exceptions techniques ingénieuses en s'appuyant sur des interprétations sophistiquées de la physique impliquée par ces processus. Mais elles n'ont pas produit beaucoup de conversions. L'embarras des dualistes sur ce point est en fait plus simple

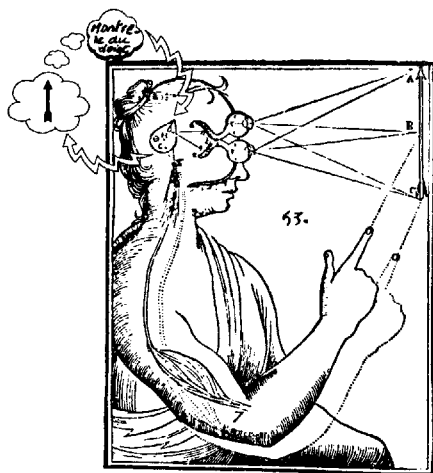


Figure 1.2

que la mention des lois présumées de la physique ne le suggère. C'est la même incohérence que celle que relèvent les enfants – mais qu'ils tolèrent avec joie pour rire – dans les histoires de Casper le gentil fantôme (fig. 1.3, p. 53). Comment Casper peut-il à la fois passer à travers les murs et attraper une serviette qui tombe ? Comment la substance mentale peut-elle à la fois échapper à toute mesure physique et contrôler le corps ? Un fantôme dans la machine ne nous est d'aucune aide pour nos théories s'il ne peut mouvoir des choses autour de lui – comme un esprit frappeur bruyant qui peut renverser une lampe ou claquer une porte. Mais toute chose qui peut mouvoir une chose physique est elle-même une chose physique (bien que ce soit peut-être une sorte de chose physique étrange et jusqu'à présent non étudiée).

Pourquoi, par conséquent, ne pas envisager une autre option. Ne peut-on conclure que la substance mentale est en fait une forme particulière de matière ? Dans les séances de spiritisme qui avaient lieu à l'époque victorienne, les médiums faisaient souvent sortir du néant quelque chose qu'ils appelaient « ectoplasme ». C'était une étrange substance évanescente qui était

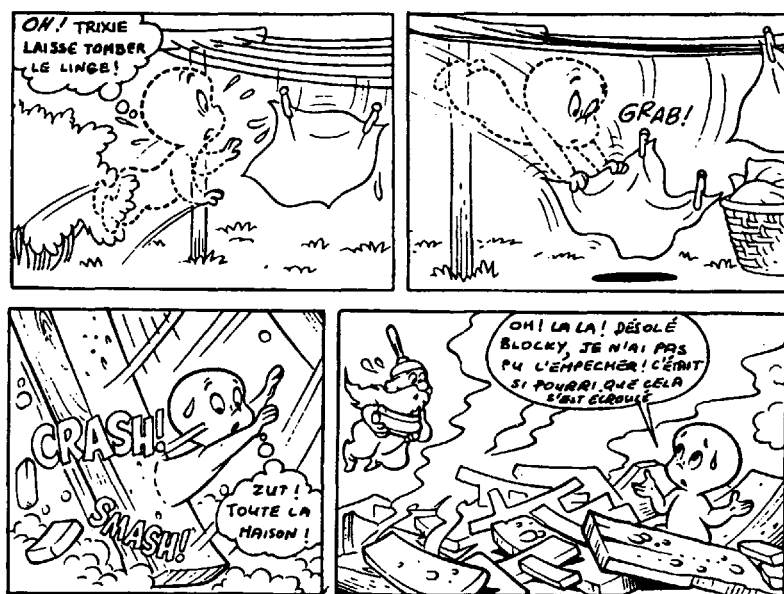


Figure 1.3

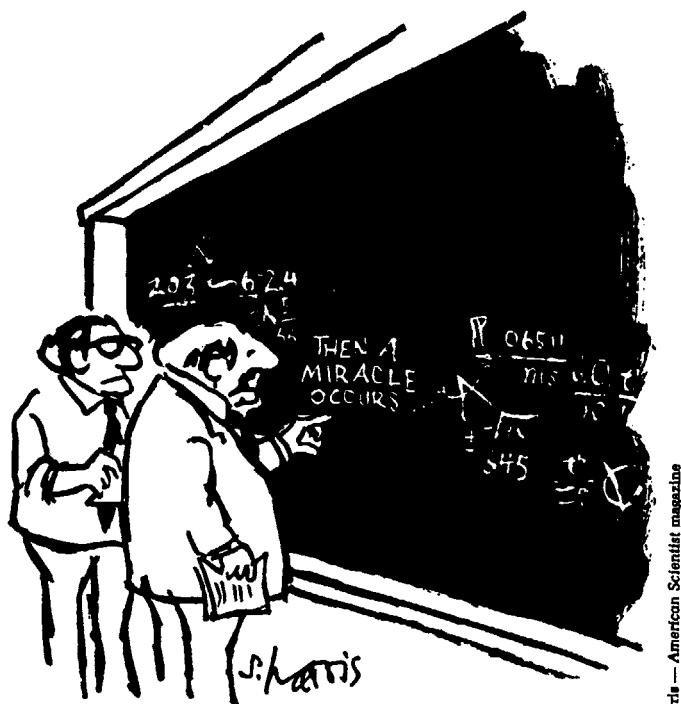
supposée constituer le matériau de base du monde des esprits, mais qui pouvait être retenue prisonnière dans un vase de verre, et qui filtrait, suintait et réfléchissait la lumière comme n'importe quelle matière ordinaire. Ces trucs frauduleux ne devraient pas nous dissuader de nous demander, plus sobrement, si la substance mentale ne pourrait pas être quelque chose qui soit au-dessus et au-delà des atomes et des molécules qui composent le cerveau, bien que ce soit cependant une forme de matière impossible à analyser scientifiquement. L'ontologie liée à une théorie est le catalogue des choses et des types de choses que la théorie réputée exister. L'ontologie des sciences physiques incluait autrefois le « calorique » (la substance dont était faite la chaleur, en fait) et « l'éther » (la substance qui remplissait l'espace et qui était le médium des vibrations de la lumière de la même façon que l'air ou l'eau peuvent être le médium des vibrations du son). On ne prend plus ces choses au sérieux, alors qu'on inclut à présent les neutrinos, l'antimatière et les trous noirs dans l'ontologie scientifique usuelle. Peut-être faut-il augmenter globalement l'ontologie des sciences physiques pour rendre compte du phénomène de la conscience.

C'est précisément le genre de révolution en physique que vient de proposer le physicien et mathématicien Roger Penrose dans *L'Esprit, les ordinateurs et les lois de la physique* (1989). Bien que je ne croie pas moi-même qu'il ait réussi à établir que cette révolution s'imposait ⁴, il est important de remarquer qu'il a pris garde de ne pas tomber dans le piège du dualisme. Quelle est la différence ? Penrose dit clairement qu'il cherche, par la révolution qu'il propose, à rendre l'esprit conscient *plus* accessible à l'enquête scientifique, et non pas moins. Ce n'est sûrement pas un hasard si les dualistes qui ont eu le courage d'avouer à tout le monde qu'ils soutiennent cette thèse ont annoncé avec candeur et naturel qu'ils n'ont pas la moindre théorie quant à la manière dont l'esprit fonctionne. Cela dépasse les pouvoirs humains, insistent-ils ⁵. On peut soupçonner en fait que la caractéristique la plus séduisante de la substance mentale est sa promesse de rester *si* mystérieuse qu'elle doit pour toujours tenir la science en échec.

Cette position fondamentalement antiscientifique adoptée par le dualisme est, à mon sens, ce qui le déconsidère le plus ; c'est la raison pour laquelle, dans ce livre, j'adopte la règle apparemment dogmatique selon laquelle on doit éviter le dualisme *à tout*

prix. Cela ne veut pas dire que je pense pouvoir donner une preuve définitive de la fausseté ou de l'incohérence du dualisme, mais plutôt que, vue la façon dont le dualisme se vautre dans son mystère, *accepter le dualisme c'est renoncer* (comme dans la figure 1.4.).

On s'accorde en général sur ce point, mais d'une façon aussi vague qu'unanime, en colmatant quelques craquelures troublantes sur le mur matérialiste. Les scientifiques et les philosophes sont peut-être parvenus à un certain consensus en faveur du matérialisme. Mais, comme nous le verrons, il est plus difficile de se débarrasser de certaines conceptions dualistes que ne l'ont cru certains matérialistes contemporains. Pour remplacer cor-



"JE PENSE QUE TU DEURAI ÊTRE PLUS EXPLICITE À L'ÉTAPE DEUX."

© 1975 Sidney Harris — American Scientist magazine

Figure 1.4

rectement les images dualistes traditionnelles, il faudra procéder à des ajustements étonnants dans nos façons de penser habituelles, et ils choqueront tout autant les intuitions des scientifiques que celles des profanes.

Je ne considère pas que le fait que ma théorie semble de prime abord entrer en conflit avec la sagesse commune soit inquiétant. Au contraire, nous ne devrions pas nous attendre à ce qu'une bonne théorie de la conscience soit facile à lire – comme si elle devait immédiatement « sonner juste » et nous faire nous exclamer, avec une sorte de fierté secrète : « Bien sûr ! Je l'ai toujours su ! C'est évident, une fois qu'on me l'a montré ! » Si l'on pouvait obtenir un jour ce genre de théorie, nous l'aurions sûrement déjà trouvée. Les mystères de l'esprit sont si anciens et nous avons fait si peu de progrès dans ce domaine qu'il est fort probable que certaines des choses que nous tendons à considérer comme évidentes ne le sont tout simplement pas. J'introduirai bientôt mes candidats. Certains chercheurs contemporains – peut-être même la majorité d'entre eux – continuent à prétendre sans broncher que, pour eux, l'esprit est tout simplement un autre organe, comme le rein ou le pancréas, qu'on ne doit décrire et expliquer que dans les termes les plus fiables des sciences biologiques et physiques. Ils ne songeraient jamais à mentionner l'esprit ou quoi que ce soit de « mental » dans leurs travaux professionnels. Pour d'autres, qui sont des chercheurs aux théories plus audacieuses, il existe un nouvel objet d'étude qu'ils appellent l'esprit/cerveau (Churchland, 1986). Ce nouveau label à succès exprime bien le matérialisme qui domine chez ces chercheurs : ils sont prêts à admettre à la face du monde et devant eux-mêmes que ce qui fait que le cerveau est particulièrement fascinant et déroutant, c'est que d'une façon ou d'une autre, il *est* l'esprit. Mais même parmi ces chercheurs, on répugne à affronter les Grands Problèmes ; on désire reporter à plus tard les questions embarrassantes que pose la nature de la conscience.

Bien que cette attitude soit fort raisonnable et qu'elle reconnaisse la valeur de la stratégie consistant à diviser pour mieux régner, elle a pour effet de déformer certains des concepts nouveaux qui ont vu le jour au sein de ce que l'on appelle aujourd'hui les *sciences cognitives*. Presque tous les chercheurs en sciences cognitives, qu'ils se considèrent eux-mêmes comme des praticiens des neurosciences, comme des psychologues ou comme des chercheurs en intelligence artificielle, tendent à reporter à plus tard

les questions que pose la conscience ; ils restreignent leur attention aux systèmes « périphériques » et « subordonnés » de l'esprit/cerveau, qui sont réputés fournir des informations et être au service d'un « centre » qu'on imagine confusément être le lieu où « la pensée consciente » et l'« expérience » se tiennent. On tend ainsi à laisser un peu trop de travail à faire « au centre », ce qui conduit les théoriciens à sous-estimer la « quantité de compréhension » que doivent accomplir les systèmes périphériques du cerveau (Dennett, 1984b).

Par exemple, les théoriciens tendent à penser que les systèmes perceptifs fournissent de l'« input » à une arène centrale de pensée, laquelle fournit en retour le « contrôle » ou la « direction » aux systèmes relativement périphériques qui gouvernent le mouvement corporel. On considère aussi que cette arène centrale se procure de l'information dans les divers systèmes de la mémoire qui sont à son service. Mais l'idée même qu'il existe des divisions théoriques importantes entre ces sous-systèmes présumés tels que « la mémoire à long terme » et le « raisonnement » (ou la « planification ») est plus un artefact lié à la stratégie du diviser-pour-mieux-régner que quoi que ce soit qu'on puisse trouver dans la nature. Comme nous le verrons bientôt, le fait de prêter attention exclusivement à des sous-systèmes spécifiques de l'esprit/cerveau conduit souvent à une sorte de myopie théorique ; elle empêche les théoriciens de voir que leurs modèles présupposent encore que quelque part, convenablement caché dans l'obscur « centre » de l'esprit/cerveau, il y a un Théâtre Cartésien, un lieu où « tout se réunit » et où se trouve la conscience. Cela peut sembler une bonne idée, une idée inévitable, mais tant que nous n'aurons pas vu, dans le détail, pourquoi ce n'est pas une bonne idée, le Théâtre Cartésien continuera à attirer des foules de théoriciens sous le charme de son illusion.

Le défi

Dans la section précédente, j'ai noté que si le dualisme est tout ce à quoi nous pouvons aboutir, alors nous ne pouvons pas comprendre la conscience humaine. Certaines personnes sont convaincues que nous ne le pouvons pas de toute façon. Un tel défaitisme, aujourd'hui, alors même qu'abondent les avancées scientifiques prêtes à être exploitées, me semble futile et même

pathétique. Mais je suppose que ce peut être la triste vérité. Peut-être la conscience ne peut-elle réellement être expliquée, mais comment le saurons-nous tant que personne n'aura essayé ? Je pense que l'on comprend déjà de nombreuses pièces du puzzle – et en fait la plupart. On a seulement besoin de les placer correctement... moyennant un petit coup de main de ma part. Ceux qui veulent défendre l'Esprit contre la Science devraient me souhaiter bonne chance dans ma tentative : s'ils ont raison, mon projet est voué à l'échec, mais si je fais mon travail à peu près correctement, mon échec devrait nous éclairer sur la raison pour laquelle la science échoue toujours. Cela leur fournira enfin leur argument contre la science, et j'aurai fait tout le sale boulot pour eux.

Les règles de base qui régissent mon projet sont simples :

1. *Pas de Tissu Miracle*. J'essaierai d'expliquer toutes les propriétés de la conscience humaine qui font problème en restant dans le cadre de la science physique contemporaine ; à aucun moment, je ne ferai appel à des forces, à des substances ou à des pouvoirs organiques inexplicables ou inconnus. En d'autres termes, j'ai l'intention de voir ce que l'on peut faire dans les limites conservatrices de la science ordinaire et de n'en appeler à une révolution dans le matérialisme qu'en dernier recours.

2. *Pas d'anesthésie feinte*. On a dit des behavioristes qu'ils feignent l'anesthésie – qu'ils prétendent ne pas avoir les expériences dont nous savons bien qu'ils les partagent avec nous. Si je veux nier l'existence d'un trait problématique quelconque de la conscience, c'est à moi qu'incombe la tâche de *montrer* qu'il est en quelque sorte illusoire.

3. *Ne chipotons pas sur les détails empiriques*. J'essaierai de rendre compte correctement de tous les faits scientifiques, dans la mesure où on les connaît aujourd'hui. Mais on débat abondamment pour savoir lesquelles, parmi les avancées excitantes, résisteront au test du temps. Si je devais me restreindre aux « faits qui sont passés dans les manuels », je serais incapable de recourir à certaines des découvertes récentes les plus révélatrices (si elles sont bien telles). Et si je me laissais guider par l'histoire récente, je *finirais* quand même par avancer sans le vouloir des données fausses. Certaines des « découvertes » concernant la vision pour lesquelles David Hubel et Torstein Wiesel ont obtenu un prix Nobel mérité en 1981 sont aujourd'hui contes-

tées ; la fameuse théorie « retinex » de la vision des couleurs, due à Edwin Land, qui a été considérée par la plupart des philosophes de l'esprit et par d'autres non-spécialistes comme un fait établi depuis plus de vingt ans, n'est de toute évidence pas aussi prisée chez les savants qui étudient la vision ⁶.

Par conséquent, puisqu'en tant que philosophe, j'entends établir *des possibilités* (et rejeter de prétendues impossibilités), je m'en tiendrai à des esquisses de théories plutôt qu'à des théories complètes, empiriquement confirmées. Une esquisse de théorie, ou un modèle de la façon dont le cerveau *pourrait* fonctionner, peut transformer une perplexité en un programme de recherche : si ce modèle ne convient pas tout à fait, certaines autres variations plus réalistes ne feront-elles pas l'affaire ? (L'explication de la production d'hallucination esquissée dans le prélude en est un exemple.) Ce genre d'esquisse est directement et explicitement vulnérable aux infirmations empiriques. Mais si vous voulez soutenir que mon esquisse n'est pas une explication *possible* d'un phénomène, vous devez montrer ce qu'il *doit* négliger ou ce qu'il *ne peut pas* faire ; si vous soutenez simplement que mon modèle pourrait bien être incorrect sur de nombreux points de détail, je m'accorderai avec vous. Ce qui cloche dans le dualisme cartésien, par exemple, ce n'est pas que Descartes a choisi la glande pinéale – par opposition, par exemple, à l'hypothalamus ou à l'amygdale – comme lieu d'interaction avec l'esprit, mais l'*idée même* qu'un tel lieu d'interaction de l'esprit avec le cerveau puisse exister. Ce que l'on tient pour du chipotage change, bien entendu, au fur et à mesure que la science avance, et différents théoriciens ont différents critères. J'essaierai d'éviter de pécher par excès de détails pour mettre en évidence le contraste avec la philosophie de l'esprit traditionnelle, mais aussi pour fournir à ceux qui me critiquent d'un point de vue empirique une cible plus nette pour leur tir.

Dans ce chapitre, nous avons rencontré les éléments fondamentaux du mystère de la conscience. Le caractère mystérieux de la conscience est l'un de ses traits principaux – peut-être même est-ce l'un des traits sans lequel elle ne peut survivre. Puisqu'on prend en général en compte cette possibilité, même si elle reste peut-être mal évaluée, la prudence penche en faveur des doctrines qui ne cherchent pas à expliquer la conscience, parce que la conscience compte beaucoup pour nous. Le dualisme,

l'idée qu'un cerveau ne peut être une chose qui pense, est tentant pour toutes sortes de raisons. Mais nous devons résister à la tentation ; « adopter » le dualisme, c'est plus ou moins accepter la défaite sans l'admettre. Adopter le matérialisme ne dissout pas par soi-même les énigmes de la conscience, pas plus que celles-ci ne tombent face aux inférences simples tirées des sciences du cerveau. D'une façon ou d'une autre, le cerveau doit être l'esprit, mais tant que l'on ne voit pas de façon détaillée comment c'est possible, notre matérialisme n'expliquera pas la conscience ; il promettra seulement de l'expliquer un beau jour. Cette promesse ne peut être tenue, ai-je suggéré, tant que nous n'apprenons pas à abandonner plus nettement l'héritage de Descartes. En même temps, quoi que nos théories matérialistes puissent aussi expliquer, elles n'expliqueront pas *la conscience*, si nous négligeons les faits relevant de l'expérience que nous connaissons si intimement « de l'intérieur ». Dans le chapitre suivant, nous allons faire l'inventaire de ces faits.

NOTES

1. Voir mon *Elbow Room* (1984), chapitre 4, où je discute ce point plus avant.

2. « *The buck stops here.* »

3. Un petit nombre d'âmes intrépides (caractérisation à laquelle ils ne peuvent s'opposer !) ont bravé le courant : le livre d'Arthur Koestler fièrement intitulé *The Ghost in the Machine* (1967) et celui de Popper et Eccles *The Self and Its Brain* (1977) ont été écrits par des auteurs incontestablement éminents. On trouvera aussi deux défenses iconoclastes et pénétrantes du dualisme dans les deux livres de Zeno Vendler, *Res Cogitans* (1967) et *The Matter of Minds* (1984).

4. Voir « Murmurs in the Cathedral » (Dennett, 1989c), mon compte rendu de ce livre.

5. Eccles a proposé l'idée que l'esprit non physique est composé de millions de « psychons » qui interagissent avec des millions de « dendrons » (des séries de cellules pyramidales) dans le cortex ; chaque psychon correspond en gros à ce que Descartes ou Hume appelleraient une idée – comme l'idée de rouge, ou l'idée de rondeur, ou de chaleur. Mais à part cette décomposition minimale, Eccles n'a rien à nous dire sur les parties, les activités, les principes d'action ou d'autres propriétés de l'esprit non physique.

6. Un exposé fascinant du statut de la théorie de Land est fourni par le philosophe C.L. Hardin dans un appendice de son livre, *Color for Philosophers ; Unwearing the Rainbow* (1988).

Chapitre 2

Une visite au jardin phénoménologique

Bienvenue au phénomène

Imaginez que nous ayons affaire à un fou qui nous dirait que les animaux n'existent pas. Nous pourrions décider de le confronter à son erreur en le conduisant au zoo. « Regarde ! Que sont ces choses, sinon des animaux ? » Nous ne nous attendrions pas à le guérir ainsi de sa maladie, mais au moins nous aurions la satisfaction de rendre claire à nos propres yeux la folie de ce qu'il avançait. Mais supposons qu'il nous dise alors : « Oh ! je sais très bien que ces choses-ci existent – les lions, les autruches et les boas constrictors. Mais qu'est-ce qui te fait penser que ces soi-disant animaux sont bien *des animaux* ? En fait, ce ne sont que des *robots* couverts de fourrure – en fait, certains sont couverts de plumes ou d'écailles. » Il se pourrait bien que ce fût de la folie, mais ce serait là une forme plus défendable de folie. Ce fou a seulement une conception révolutionnaire de la nature ultime des animaux ¹.

Les zoologues sont experts quant à la nature ultime des animaux et les jardins zoologiques – les zoos, en bref – sont au service du projet qui consiste à familiariser le bon peuple avec les objets de leur savoir. Si les zoologues devaient découvrir que notre fou avait (passez-moi l'expression) raison, ils découvriraient comment bien utiliser leur zoo en essayant d'expliquer leur découverte. Ils pourraient dire : « Il appert que les animaux – vous savez : ces objets familiers que nous gardons au zoo – ne sont pas ce que nous croyions qu'ils étaient. Ils sont si différents de ce que nous pensions que nous ne devrions pas les appeler

des animaux. C'est pourquoi, voyez-vous, en fait, *il n'y a pas* d'animaux, au sens usuel de ce terme. »

Les philosophes et les psychologues utilisent souvent le terme *phénoménologie* pour désigner génériquement toutes les choses — la faune et la flore, pourrait-on dire — qui habitent le monde de notre expérience consciente : les pensées, les odeurs, les déman-gaisons, les douleurs, les vaches violettes dans notre imagination. Plusieurs sources distinctes sont à l'origine de cet usage. Au xvii^e siècle, Kant distinguait les « phénomènes », les choses qui apparaissent, des « noumènes », les choses telles qu'elles sont en soi, et au ix^e siècle, alors que se développaient les sciences naturelles ou physiques, le terme *phénoménologie* en est venu à désigner l'étude simplement descriptive d'un sujet quelconque, pris d'un point de vue neutre ou préthéorique. La phénoménologie du magnétisme, par exemple, a d'abord été étudiée au xvi^e siècle par William Gilbert, mais l'explication de cette phénoménologie a dû attendre la découverte des relations entre magnétisme et électricité au xix^e siècle et le travail théorique de Faraday, Maxwell et autres. C'est par allusion à cette distinction entre l'observation rigoureuse et l'explication théorique que l'école ou le courant philosophique connu sous le nom de Phénoménologie (avec un P majuscule) a pris son essor au début du xx^e siècle avec l'œuvre d'Edmund Husserl. Elle avait pour but de donner un nouveau fondement à l'ensemble de la philosophie (et en fait à la connaissance tout entière) en s'appuyant sur une technique particulière d'introspection, grâce à laquelle le monde extérieur et toutes ses implications étaient supposés être « mis entre parenthèses » par un acte particulier de l'esprit appelé *épochè*. Cette procédure était réputée produire un état d'esprit et une forme d'enquête par lesquels le Phénoménologue était supposé obtenir une connaissance intime des objets purs de l'expérience consciente, appelés *noemata*, dégagés des distorsions et des modifications qu'ils subissent du fait de nos théories et de nos pratiques courantes. À l'instar des autres tentatives pour évacuer les interprétations et révéler les faits fondamentaux de la conscience à l'observation rigoureuse, comme le mouvement impressionniste en art et les psychologies introspectives de Wundt, Titchener et d'autres, la Phénoménologie n'a pas réussi à découvrir une méthode unique et bien établie sur laquelle tout le monde puisse se mettre d'accord.

Par conséquent, bien qu'il y ait des zoologues, il n'y a pas

réellement de phénoménologues : il n'y a pas d'experts incontestables quant à la nature des choses qui nagent dans le courant de la conscience. Mais nous pouvons suivre l'usage récent et adopter ce terme (avec un p minuscule) pour désigner en général les différentes entités qui apparaissent dans l'expérience consciente et dont on cherche à expliquer la nature.

J'ai jadis publié un article intitulé « Sur l'absence de phénoménologie » (1979), qui essayait de défendre le second type de folie : les choses dont la conscience est composée sont si différentes de ce que l'on croit que nous devrions cesser d'utiliser l'ancienne terminologie. Mais cette suggestion a paru si choquante à certains (« Comment pourrions-nous nous tromper sur la nature de notre vie intime ? ») qu'ils ont essayé de rejeter ma conception en l'assimilant à la première forme de folie (« Dennett ne pense pas qu'il existe des douleurs, des arômes ou des rêveries ! »). C'était là une caricature de ma conception, bien entendu, mais cette réaction était tentante. Mon problème était que je n'avais pas à ma disposition de jardin phénoménologique – disons un phénomène – pour produire mes explications. Je voulais dire : « Il appert que les choses qui nagent dans le courant de la conscience – vous savez : les douleurs, les arômes, les rêveries, les images mentales et les poussées de colère et de désir, les habitants usuels du phénomène – ces choses ne sont pas ce que nous croyions qu'elles étaient. Elles sont en fait si différentes que nous devrions les appeler autrement. »

Faisons donc rapidement le tour du jardin phénoménologique, juste pour nous assurer que nous savons de quoi nous sommes en train de parler (même si nous ne savons pas quelle est la nature ultime de ces choses). Ce sera un parcours introductif délibérément superficiel, destiné à montrer les choses à la ronde au moyen de quelques mots d'explication, en posant quelques questions, avant de nous livrer à une investigation théorique plus sérieuse dans le reste du livre. Comme je me mettrai bientôt à contester radicalement notre forme de pensée usuelle, je ne voudrais pas qu'on s'imagine que j'ignore tout simplement les choses merveilleuses qui habitent les esprits *des autres gens*.

Notre phénomène se divise en trois parties : (1) *des expériences du monde « extérieur »*, telles que des visions, des sensations de chaleur ou de froid, et des positions de nos membres ; (2) *des expériences du monde purement interne*, telles que des images produites dans l'imagination, les visions internes et les sons qui

se produisent quand vous vous abandonnez à la rêverie et quand vous parlez à votre bonnet, des souvenirs, des idées soudaines ou des pressentiments ; et (3) *des expériences d'émotions ou d'« affects »* (pour utiliser le terme bizarre qu'aiment employer les psychologues), qui vont des douleurs physiques, des chatouillements et des « sensations » de faim et de soif, en passant par les poussées émotionnelles de colère, de joie, de haine, d'embarras, de désir, d'étonnement, aux visitations moins physiques de l'orgueil, de l'anxiété, du regret, du détachement ironique, de la terreur et de la froideur.

Je ne cherche pas à justifier cette division tripartite entre l'externe, l'interne et l'affectif. Tout comme la ménagerie qui met les chauves-souris avec les oiseaux et les dauphins avec les poissons, cette taxinomie repose plus sur des similarités superficielles et sur une tradition douteuse que sur la parenté profonde entre les phénomènes, mais il nous faut bien commencer quelque part, et toute taxinomie qui nous donne quelque prise nous évitera de négliger totalement certaines espèces.

Notre expérience du monde extérieur

Commençons par les plus élémentaires de nos sens externes, le goût et l'odorat. Comme le savent la plupart des gens, nos papilles du goût ne sont en fait sensibles qu'à ce qui est sucré, acide, salé et amer ; pour l'essentiel nous « goûtons avec notre nez ». C'est pourquoi la nourriture perd sa saveur quand nous avons un rhume de cerveau. L'épithélium nasal est à l'odorat, au sens de l'odeur, ce que la rétine est au sens de la vision. Les cellules épithéliales individuelles sont très variées, chacune étant sensible à un type différent de molécule apportée par l'atmosphère extérieure. En dernière instance, c'est la *forme* de la molécule qui compte. Les molécules flottent dans le nez, comme autant de clefs microscopiques, reposant sur des cellules particulières de l'épithélium. Les molécules peuvent souvent être détectées par des concentrations étonnamment faibles de quelques parties par milliards. D'autres animaux ont une olfaction largement supérieure à la nôtre, non seulement par leur capacité à discerner plus d'odeurs, à partir de traces plus faibles (le limier est l'exemple type), mais aussi de par la plus grande finesse de leur détection temporelle et spatiale des odeurs. Nous pouvons sentir la présence

dans une pièce d'un petit amas de molécules de formaldéhyde, mais si nous parvenons à le faire, nous n'avons pas à sentir que l'amas de molécules en question a la forme d'une chaîne ou d'une région comportant des molécules perceptibles individuellement et flottant dans la pièce ; l'ensemble de la pièce, ou tout au moins un coin tout entier de la pièce, nous paraîtra rempli de l'odeur en question. L'explication de ce fait n'a rien de mystérieux : les molécules *errent* plus ou moins au sein de nos voies nasales, et leur arrivée en certains points de l'épithélium ne fournit qu'une information partielle sur l'endroit d'où elles viennent dans notre environnement, à la différence des photons qui se déplacent par lignes droites optiques à travers l'orifice minuscule qu'est l'iris, atteignant une adresse rétinale qui se projette géométriquement sur une source externe ou sur un chemin source. Si la finesse de résolution de notre vision était aussi pauvre que celle de notre olfaction, un oiseau qui volerait au-dessus de nous dans le ciel produirait en nous une vague *sensation oisiforme* pendant un instant. (La vision de certaines espèces est d'une pauvreté similaire : leur capacité de résolution et de discrimination n'est pas meilleure. Mais l'effet que cela produit pour l'animal d'avoir une vue aussi défectueuse est une autre affaire ; nous y reviendrons dans un chapitre ultérieur.)

Nos sens du goût et de l'odorat sont phénoménologiquement attelés l'un à l'autre, de même que nos sens du toucher et de la kinesthésie, et que le sens de la position et du mouvement de nos membres et d'autres parties du corps. Nous « sentons » les choses en les touchant, en les attrapant, en les poussant de multiples manières ; mais les sensations conscientes qui en résultent, bien qu'elles paraissent à un regard naïf être de simples « traductions » de la stimulation des récepteurs du toucher sous la peau, sont, une fois encore, les produits d'un processus élaboré d'intégration de l'information à partir d'un ensemble de sources variées. Bandez vos yeux et prenez un bâton (ou un crayon ou un stylo) dans votre main. Touchez diverses choses autour de vous avec cette baguette ; vous pouvez reconnaître leur texture sans effort, comme si votre système nerveux avait des récepteurs sensoriels à l'extrémité de la baguette. Un effort particulier, et dans l'ensemble sans résultat, est nécessaire pour être attentif à la manière dont on ressent le bâton au bout de ses doigts, pour sentir la manière dont il vibre ou résiste au mouvement qu'on lui imprime quand il entre en contact avec les diverses surfaces.

Ces transactions entre le bâton et les récepteurs du toucher sous la peau (aidés dans la plupart des cas par des sons à peine perceptibles) fournissent l'information que votre cerveau intègre dans une reconnaissance consciente de la texture du papier, du carton, de la laine ou du verre ; mais ces processus compliqués d'intégration sont tout sauf transparents à la conscience. C'est-à-dire que nous ne savons pas – et ne pouvons pas savoir – comment « nous » le faisons. Pour prendre un cas encore plus indirect, songez à la manière dont vous pouvez sentir la viscosité d'une tache d'huile sur la route sous les roues de votre voiture quand vous tournez à un coin de rue. Le point focal phénoménologique de contact est celui où le caoutchouc rencontre la route ; ce n'est pas un point de votre corps innervé, assis, habillé, sur le siège de l'auto, ou de vos mains gantées sur le volant.

À présent, toujours les yeux bandés, abaissez votre baguette. Demandez à quelqu'un de vous présenter un morceau de porcelaine chinoise, un morceau de plastique et des morceaux de bois et de métal poli. Ces matériaux sont tous très lisses et glissants, et pourtant vous aurez peu de difficultés à discerner la qualité particulière de leur texture lisse. Ce n'est pas parce que vous avez des récepteurs à porcelaine chinoise. La différence de conductivité de la chaleur entre chaque substance est apparemment le facteur le plus important, mais il n'est pas essentiel : vous pouvez parfois être surpris de constater combien il vous est facile de distinguer ces surfaces les unes des autres par « sensation » en n'utilisant que la baguette. Ces succès doivent dépendre des vibrations ressenties à travers la baguette ou de différences indescriptibles – bien que détectables – dans les claquements et les bruits entendus quand elle érafle les surfaces. Mais on a *l'impression* que tout se passe comme si nos terminaisons nerveuses étaient dans la baguette, car vous *sentez* les différences sur les surfaces *au bout* de la baguette.

Ensuite, considérons l'ouïe. La phénoménologie de l'ouïe consiste en toutes les sortes de sons que nous pouvons entendre : de la musique, des mots parlés, des explosions et des coups de sifflets, des sirènes, des gazouillis et des cliquètements. Les théoriciens de l'ouïe sont souvent tentés « de mettre en marche le petit orchestre dans notre tête ». C'est une erreur, et pour être sûrs que nous l'identifions et l'évitons, je vais raconter une fable.

Il était une fois, à peu près au milieu du XIX^e siècle, un inven-

teur au regard halluciné qui se mit à débattre avec un philosophe obstiné, Philéas. L'inventeur avait annoncé qu'il avait l'intention de construire un dispositif qui pourrait automatiquement « enregistrer » et ensuite « rejouer » avec une « fidélité » parfaite un orchestre et un chœur jouant la *Neuvième Symphonie* de Beethoven. « C'est un non-sens, dit Philéas. C'est strictement impossible. Je peux aisément imaginer un dispositif mécanique qui enregistrerait une suite de touches de piano qu'on frapperait, puis qui contrôlerait la reproduction de cette séquence sur un piano préparé à cet effet. On pourrait le faire par exemple avec un rouleau de papier perforé. Mais songez à l'immense variété de sons et à leurs modes de production quand il s'agit de rendre la *Neuvième* de Beethoven ! Il y a des centaines de voix humaines de hauteur et de timbres différents, des douzaines de cordes maniées par l'archet, les cuivres, les instruments à vent, les percussions. Le dispositif qui pourrait rejouer ensemble une telle variété de sons serait tellement encombrant et monstrueux qu'à côté, le plus énorme des orgues aurait l'air minuscule. S'il était joué avec la "haute fidélité" que vous proposez, il devrait sans aucun doute incorporer littéralement une équipe d'esclaves humains pour se charger des parties vocales, et ce que vous appelez "l'enregistrement" d'une exécution particulière avec toutes les nuances devrait comporter des centaines de partitions – une pour chaque musicien – avec des milliers, voire des millions d'annotations. »

L'argumentation de Philéas est toujours étrangement convaincante : il est stupéfiant que tous ces sons puissent se surimposer fidèlement par l'intermédiaire d'une transformation de Fourier en une unique ligne ondulante ciselée dans un long disque ou représentée magnétiquement sur une bande magnétique ou optiquement sur la bande-son d'un film. Il est encore plus stupéfiant qu'un unique cône de papier, qu'un électro-aimant guidé par cette ligne ondulante unique fait osciller en cadence, puisse faire justice autant à la sonnerie d'une trompette, qu'au raclement des cordes d'un banjo, à la voix humaine et au son d'une bouteille de vin pleine se brisant sur le trottoir. Philéas ne pouvait pas imaginer un dispositif si puissant ; il confondit son manque d'imagination avec une intuition profonde sur ce qui est nécessaire.

La « magie » des transformations de Fourier ouvre à notre pensée une nouvelle gamme de possibilités. Mais nous devons

noter qu'elle n'élimine pas par elle-même le problème qui troublait l'esprit de Philéas ; elle ne fait que le déplacer. Car bien que nous puissions, nous qui sommes sophistiqués, rire de Philéas ne parvenant pas à comprendre comment la structure de compression et de raréfaction de l'air qui stimule l'oreille peut être enregistrée et reproduite, nos sourires condescendants quitteront nos lèvres quand nous considérerons la question suivante : qu'arrive-t-il au signal à partir du moment où l'oreille l'a bien reçu ?

À partir de l'oreille, une barrière supplémentaire de séries de signaux modulés – à présent plus ou moins analysés et décomposés en courants parallèles, rappelant étrangement les centaines de partitions dont parlait Philéas – progressent vers le centre sombre du cerveau. Ces séries de signaux ne sont pas plus des *sons entendus* que ne le sont les lignes ondulantes sur le disque ; ce sont des suites de pulsations électrochimiques qui naviguent vers les axones des neurones. Y a-t-il un lieu encore plus central dans le cerveau où ces séries de signaux contrôlent les actions de l'orgue qui se trouve dans l'immense théâtre de l'esprit ? Quand, après tout, ces signaux sans tonalité sont-ils finalement *traduits* en sons subjectivement entendus ?

Nous ne voulons pas rechercher des lieux dans le cerveau qui vibreraient comme des cordes de guitare, pas plus que nous ne voulons trouver des lieux dans le cerveau qui deviendraient violets quand nous imaginons une vache violette. Ce sont là manifestations des impasses, ou ce que Gilbert Ryle (1949) appellerait des erreurs de catégorie. Mais alors que *pourrions-nous* trouver dans le cerveau qui nous permettrait d'avoir la satisfaction que nous avons atteint le fin mot de l'histoire quant à l'expérience auditive ² ? Comment des propriétés physiques complexes des événements se produisant dans le cerveau pourraient-elles constituer les propriétés fantastiques des sons que nous entendons – ou même seulement en rendre compte ?

De prime abord, ces propriétés paraissent inanalysables – ou pour utiliser un adjectif qu'aiment bien les phénoménologues – *ineffables*. Mais au moins certaines de ces propriétés apparemment atomiques et homogènes peuvent donner lieu à la découverte des éléments qui les composent et se prêter à une description. Prenez une guitare et branchez la note mi basse ou majeure (sans presser la moindre touchette). Écoutez attentivement le son produit. A-t-il des composants qu'il est possible de

décrire ou bien forme-t-il un tout rendant ineffablement une sonorité de guitare ? Beaucoup de gens choisiront de décrire leur phénoménologie de cette seconde manière. À présent, tirez à nouveau sur la corde ouverte et faites glisser doucement votre doigt vers la touchette de l'octave pour créer un « harmonique » élevé. Soudain, vous entendez un nouveau son : d'une certaine façon, il est plus « pur » et bien sûr d'un octave plus élevé. Certaines personnes disent que c'est un son entièrement nouveau, alors que d'autres décrivent leur expérience en disant que « c'est le bas de la note qui est tombé » – en laissant seulement le haut. Tirez alors la corde ouverte une nouvelle fois. Cette fois-ci, vous pouvez entendre, de façon étonnamment distincte, le ton harmonique qui était isolé quand vous avez tiré une seconde fois la corde. L'homogénéité et l'ineffabilité de la première expérience ont disparu ; elles sont remplacées par une dualité qui peut être aussi directement appréhendée et clairement décrite que n'importe quel accord.

La différence entre ces deux expériences est frappante ; mais la complexité nouvelle ressentie au moment de tirer sur la corde pour la troisième fois était *là* d'entrée de jeu (perçue et discriminée). Certains travaux ont pu montrer que c'est seulement par l'intermédiaire de la structure harmonique complexe que vous pouvez être capable de reconnaître le son comme étant celui d'une guitare plutôt que d'un luth ou d'une harpe. Ces travaux peuvent nous aider à *rendre compte* des différentes propriétés des expériences auditives, en analysant les composants informationnels et les processus qui les synthétisent, en nous permettant de prédire et même de provoquer synthétiquement des expériences auditives particulières. Mais ces travaux semblent laisser toujours ouverte la question de savoir de quoi ces propriétés sont *constituées*. Pourquoi la trame d'harmoniques créée par la guitare sonne-t-elle *comme ceci* et la trame provenant d'un luth *comme cela* ? Nous n'avons toujours pas répondu à cette question résiduelle, même si nous lui avons ôté de son tranchant en montrant qu'au moins, certaines propriétés considérées au départ comme ineffables se prêtent en définitive à une certaine analyse et à une certaine description ³.

Les recherches sur la perception auditive suggèrent qu'il existe des mécanismes spécialisés qui déchiffrent différentes sortes de sons, un peu comme les composants imaginaires de la machine à rejouer la musique de Philéas. Les sons émis par la voix, en particulier, semblent relever de ce qu'un ingénieur appellerait

des mécanismes spéciaux. La phénoménologie de la perception du son suggère qu'il existe un mécanisme du cerveau qui restructure entièrement les informations d'entrée. C'est un mécanisme un peu semblable au studio d'enregistrement où l'on mixe de multiples canaux d'enregistrement, où on les améliore et où on les ajuste de diverses manières pour créer le « moule » stéréo de base à partir duquel les enregistrements ultérieurs sont copiés sur divers supports.

Par exemple, nous entendons la voix dans notre langue maternelle comme une suite de mots distincts séparés par des espaces de silence. C'est-à-dire que nous avons un sens net des frontières entre les mots, qui ne peuvent pas être composées de bords ou de lignes de couleur et qui ne semblent pas marquées par des bips ou des déclics. Que peuvent bien être ces frontières sinon des espaces de silence de diverses durées – comme les espaces qui séparent les lettres et les mots dans le code Morse ? Quand des expérimentateurs demandent à des sujets de noter de diverses manières ces espaces entre les mots et de les relever, ils n'ont aucune difficulté à le faire. Il semble bien y avoir des lacunes. Mais si l'on regarde le profil acoustique d'énergie du signal, les régions où l'énergie est la plus faible (les moments qui s'approchent le plus du silence) ne s'accordent pas bien avec les frontières des mots. La segmentation des sons du discours parlé est un processus qui impose des frontières fondées sur la structure grammaticale du langage et non sur la structure physique de l'onde acoustique (Liberman et Studdert-Kennedy, 1977). Cela nous aide à comprendre pourquoi le discours parlé qui provient d'une autre langue nous paraît un torrent embrouillé et non segmenté de sons : les mécanismes spéciaux dans le « studio d'enregistrement » du cerveau n'ont pas la trame grammaticale nécessaire à la segmentation appropriée, en sorte que le mieux qu'ils puissent faire est de passer une version du signal reçu, essentiellement non retouché.

Quand nous percevons le son d'une voix, nous avons conscience de quelque chose de plus que des simples identités et des catégories grammaticales des mots. (Si c'était tout ce dont nous étions conscients, nous ne serions pas capables de déterminer si nous sommes en train d'entendre ou de lire les mots.) Les mots sont clairement démarqués, ordonnés et identifiés, mais ils sont aussi revêtus de diverses propriétés sensibles. Par exemple, je viens juste d'entendre la voix distinctivement britannique de mon ami

Nick Humphrey, qui me met gentiment en boîte, sans vraiment se moquer de moi. *J'entends* son sourire, me semble-t-il, et dans mon expérience se trouve inclus le sentiment que le rire était derrière les mots, prêt à sortir comme le soleil d'un écran de nuages. Les propriétés dont nous sommes conscients ne sont pas seulement la montée et la descente de l'intonation ; ce sont aussi les grincements, les soufflements et les zézaiements de la voix, pour ne rien dire des geignardises, des trémolos de crainte ou des expressions plates liées à la dépression. Comme nous venons de l'observer dans le cas de la guitare, des propriétés qui paraissent au départ entièrement atomiques et homogènes se prêtent souvent à l'analyse pour peu qu'on les isole et qu'on les répète expérimentalement. Nous reconnaissons sans effort le son interrogatif d'une question – et la différence entre un son interrogatif britannique et un son interrogatif américain. Mais il nous faut plusieurs expérimentations sur la variation d'un thème pour pouvoir décrire avec certitude et précision les différences entre les contours des intonations qui produisent ces différentes saveurs auditives.

« Saveurs » semble être la métaphore appropriée dans ce cas, sans aucun doute, parce que notre capacité à analyser les saveurs est limitée. Les preuves bien connues mais toujours surprenantes du fait que nous goûtons avec notre nez montrent que nos pouvoirs de goût et d'olfaction sont si rudimentaires que nous avons du mal à identifier ne serait-ce que les canaux par lesquels passe l'information reçue. Cette opacité ne se réduit pas au goût et à l'odorat : quand nous entendons des tons de fréquence très basse – comme les notes les plus basses et les plus profondes jouées par l'orgue d'une église – ils semblent plus produits par notre sentiment des vibrations de notre corps que par l'attention prêtée aux vibrations de nos oreilles. Nous sommes surpris d'apprendre qu'un certain fa dièse, « exactement deux octaves au-dessous du plus bas fa dièse que je puisse chanter », peut en fait être *entendu* avec le fond de mon pantalon plutôt qu'avec mes oreilles.

Pour finir, tournons-nous brièvement vers la vision. Quand nos yeux sont ouverts, nous avons le sentiment d'être en face d'un vaste champ, souvent appelé champ phénoménal ou champ visuel, dans lequel les choses apparaissent, colorées et situées à diverses profondeurs ou distances par rapport à nous, en mouvement ou en repos. Nous considérons naïvement que la plupart des traits dont nous avons l'expérience sont des propriétés objec-

tives des choses extérieures, observées « directement » par nous. Mais dès notre enfance, nous reconnaissons vite l'existence d'une catégorie intermédiaire d'entités – les lumières aveuglantes, les éclairs, les chatoiements, les bords flous – dont nous savons qu'elles sont d'une certaine façon le produit des interactions entre les objets, la lumière et notre équipement visuel. Nous continuons à voir ces entités comme étant « là devant » plutôt qu'en nous, à quelques exceptions près : la douleur éprouvée à regarder directement le soleil ou une lumière qui se met soudain à briller quand nos yeux sont adaptés au noir, ou l'oscillation du champ phénoménal qui nous donne la nausée quand nous sommes pris de vertige. Il vaudrait mieux décrire ces phénomènes comme des « sensations dans les yeux » : elles sont plus proches des pressions et des picotements que nous éprouvons quand nous nous frottons les yeux que des propriétés externes qui caractérisent en temps normal les choses vues.

Parmi les choses extérieures qu'on peut voir dans le monde physique, on trouve les images. Ce sont des exemples si paradigmatiques de choses-qui-sont-là-pour-être-vues que nous tendons à oublier qu'elles sont une addition plutôt récente à l'environnement visible, vieilles seulement de quelques dizaines de milliers d'années. Grâce à l'art humain récent et à l'artifice, nous sommes aujourd'hui entourés d'images, de cartes, de diagrammes, à la fois immobiles et en mouvement. Ces images physiques, qui ne sont qu'une sorte de « matériau brut » pour les processus de la perception visuelle, sont devenues un modèle quasiment irrésistible de « produit final » de la perception visuelle : des « images dans la tête ». Nous sommes enclins à dire : « Bien sûr que le produit de la vision est une image dans l'esprit [ou dans le cerveau]. Que pourrait-il être d'autre ? Certainement pas un air de musique ou une saveur ! » Nous allons traiter ces maladies curieuses et si répandues de l'imagination de bien des façons avant de nous en débarrasser, mais nous pouvons commencer par nous rappeler quelque chose : ouvrir des galeries de tableaux pour les aveugles ne sert à rien ; de même, avoir des images dans la tête demande des yeux dans la tête pour les regarder [il faudra aussi un bon éclairage !]. À supposer qu'il y ait des yeux de l'esprit dans la tête pour regarder les images dans la tête, que ferons-nous des images dans la tête de la tête produites en retour par ces yeux internes ? Comment éviter une régression à l'infini d'images et de spectateurs ? Nous ne pouvons

interrompre la régression qu'en découvrant un spectateur dont la perception parvient à ne pas créer de nouvelles images qui auraient à nouveau besoin d'un spectateur. Peut-être la régression doit-elle être bloquée dès la première étape ?

Heureusement, il existe d'autres raisons de douter du fait que la vision fonctionne d'après la théorie des images-dans-la-tête. Si la vision mettait en jeu des images dans la tête avec lesquelles nous (nos moi intérieurs) serions intimement en présence, ne serait-il pas plus aisé de dessiner des images ? Rappelons-nous combien il est difficile de dessiner une image réaliste d'une rose dans un vase, par exemple. Voici la rose aussi grosse que la vie à un mètre environ devant vous – à gauche, par exemple, de votre bloc de papier. [Je voudrais vraiment que vous imaginiez cela soigneusement.] Tous les détails visibles de la vraie rose sont nets et précis, intimement accessibles à votre regard, semble-t-il. Et pourtant, le processus apparemment si simple qui consiste à restituer une copie en noir et blanc, à deux dimensions, dans tous ces détails quelques degrés vers la droite est si compliqué que la plupart des gens abandonnent et décident qu'ils ne savent pas dessiner. Les gens trouvent qu'il est particulièrement difficile de traduire trois dimensions en deux, ce qui est assez surprenant, compte tenu du fait que ce qui semble de prime abord être la traduction inverse – voir une image réelle à deux dimensions *comme si* elle était une situation ou un objet à trois dimensions – ne demande ni effort ni attention. En fait, c'est la difficulté même que nous avons à supprimer cette interprétation inverse qui fait du processus même de dessiner une simple ligne une tâche si ardue.

Ce n'est pas seulement affaire de « coordination de la main et de l'œil ». Car des gens qui sont capables de faire de la broderie ou d'assembler des montres de poche avec une grande dextérité peuvent toujours être désespérément incapables de copier des dessins. On peut dire que c'est plus affaire de coordination *œil-cerveau*. Ceux qui maîtrisent l'art du dessin savent qu'il requiert des habitudes spéciales d'attention, des techniques telles que la capacité à tourner légèrement les yeux pour permettre de supprimer en quelque sorte la contribution de ce que l'on *sait* (la forme de la pièce de monnaie est circulaire, le dessus de la table rectangulaire) afin de pouvoir observer les angles réels sous-tendus par les lignes dans le dessin (la forme de la pièce de monnaie est elliptique, le dessus de la table trapézoïdal). Cela

aide quelquefois de surimposer une grille imaginaire verticale ou horizontale ou une paire de fils croisés dans un viseur, pour pouvoir mieux juger les angles réels des choses que l'on voit. Apprendre à dessiner consiste principalement à apprendre à dépasser les processus ordinaires de la vision de manière à rendre son expérience des choses vues dans le monde *plus proche de l'expérience de voir une image*. Cela ne peut jamais être tout à fait la même chose que voir une image, mais une fois que la chose s'est ainsi transformée, on peut, moyennant certaines autres techniques usuelles, « copier » plus ou moins ce que dont on fait l'expérience sur le papier.

Nous avons l'impression spontanée que le champ visuel est uniformément détaillé et dirigé à partir du centre vers les limites. Une expérience simple nous montre qu'il n'en est rien. Prenez un jeu de cartes à jouer et tirez-en une carte, face cachée, en sorte que vous ne puissiez voir de laquelle il s'agit. Tenez-la à bout de bras à la périphérie gauche ou droite de votre champ visuel, et tournez sa face vers vous, en prenant soin de continuer à regarder droit devant vous (par exemple, en fixant un point). Vous vous apercevrez que vous ne pouvez même pas dire si elle est rouge ou noire, ou si c'est une figure. Remarquez pourtant que vous avez nettement conscience de tout tremblement ou mouvement de la carte. Vous voyez le mouvement sans être capable de voir la forme ou la couleur de la chose qui se meut. À présent, commencez à mouvoir la carte vers le centre de votre champ visuel, toujours en faisant bien attention de ne pas déplacer votre regard. À partir de quel point pouvez-vous identifier la couleur ? À quel point la valeur et le nombre ? Notez que vous pouvez dire si c'est une figure bien avant que vous puissiez dire si c'est un valet, une reine, ou un roi. Vous serez probablement surpris de constater combien il vous est possible de mouvoir la carte près du centre du champ avant d'être capable de l'identifier.

En général, nous ne relevons pas ce défaut choquant de notre vision périphérique (de toute notre vision mis à part deux ou trois degrés autour du centre mort du champ) parce que nos yeux, à la différence des caméras de télévision, ne sont pas calmement habitués au monde ; ils oscillent sans cesse, sans que nous nous en rendions compte, dans un incessant jeu de démarquage visuel des objets auxquels nous nous intéressons quand ils viennent à occuper notre champ de vision. En suivant doucement

les choses à la trace ou en sautant en *saccades*, nos yeux fournissent à nos cerveaux des informations très fines sur tout ce qui occupe momentanément l'aire centrale fovéale du champ rétinien. (La fovéa de l'œil a une capacité de discrimination environ dix fois supérieure à celle des autres régions de la rétine.)

Notre phénoménologie visuelle, les *contenus* de notre expérience visuelle, sont très différents de tout autre mode de représentation, qu'il s'agisse des images, des films, des cartes, des phrases, des modèles réduits ou des diagrammes. Considérez ce qui est présent dans votre expérience quand vous regardez dans un stade la foule grouillante de milliers de spectateurs. Les individus sont bien trop loin de vous pour que vous puissiez les identifier, à moins que vous ne soyez aidé par une propriété saillante (le président — oui, vous pouvez dire que c'est vraiment lui, en personne ; c'est celui que vous distinguez au centre des banderoles rouge, blanc et bleu). Vous pouvez dire, visuellement, que la foule est composée d'êtres humains en raison de la façon visiblement humaine dont ils se meuvent. Il y a quelque chose de global dans votre expérience visuelle de la foule. (Il y a un effet de foule là-bas, tout comme un groupe d'arbres vus à travers une fenêtre a des allures d'ormes ou comme un parquet a l'air poussiéreux.) Mais vous ne voyez pas une grosse tache marquée « foule » ; vous voyez — d'un coup — des milliers de détails particuliers : des chapeaux rouges qui se balancent et des verres de lunettes qui scintillent, des bouts de manteaux bleus, des programmes brandis en l'air, et des points levés. Si nous essayons de peindre ce que vous voyez à la manière « impressionniste », la cacophonie de taches de couleurs n'en saisirait *pas* le contenu ; vous n'avez pas l'expérience d'une cacophonie de taches de couleurs, pas plus que vous n'avez l'expérience d'une ellipse quand vous contemplez une pièce de monnaie obliquement. Les tableaux — les images en couleurs à deux dimensions — peuvent en gros s'approcher de l'information reçue sur la rétine d'une scène à trois dimensions, et par conséquent créer en vous une impression similaire à l'impression visuelle que vous avez quand vous regardez la scène en question. Mais dans ce cas, le tableau n'est pas le tableau *de* l'impression éprouvée ; c'est plutôt quelque chose qui peut provoquer ou stimuler cette impression.

On ne peut pas plus peindre une image réaliste d'un phénomène visuel qu'on ne peut en peindre une de la justice, de la mélodie ou du bonheur. Et pourtant, il nous semble souvent

approprié, et nous y sommes irrésistiblement poussés, de parler de nos expériences visuelles comme si c'étaient des images dans l'esprit. Cela fait partie de notre phénoménologie visuelle, et par conséquent de ce que nous devons expliquer dans les chapitres suivants.

Notre expérience du monde interne

Quels sont les « matériaux bruts » de notre vie intérieure, et que faisons-nous avec eux ? La réponse ne devrait pas être difficile à découvrir, on peut supposer que nous nous contentons de « regarder et de voir », puis de noter les résultats.

Selon la tradition toujours solide des empiristes britanniques Locke, Berkeley et Hume, les sens sont les portails d'entrée des meubles de l'esprit ; une fois bien en sécurité à l'intérieur, ces matériaux peuvent être manipulés et combinés *ad libitum* pour créer un monde intérieur d'objets imaginés. Vous imaginez une vache violette qui vole en prenant le violet que vous avez retenu en voyant une grappe de raisin, les ailes que vous avez retenues en voyant un aigle et en les attachant à la vache que vous avez retenue en voyant une vache. Cette conception ne peut pas être tout à fait correcte. Ce qui entre dans l'œil, c'est une radiation électromagnétique. Elle ne devient *pas* par là même utilisable comme peuvent l'être diverses nuances de couleur quand nous voulons peindre des vaches imaginaires. Nos organes sensoriels sont bombardés d'énergie physique sous diverses formes, laquelle est « transduite » au point de contact en impulsions nerveuses qui voyagent ensuite à l'intérieur vers le cerveau. Seule de l'information passe de l'extérieur à l'intérieur, et bien que la réception de l'information puisse *provoquer* la création d'une entité phénoménologique (pour parler de façon aussi neutre que possible), il est difficile de croire que l'information elle-même – qui n'est qu'une abstraction rendue concrète dans un milieu physique modulé – pourrait *être* elle-même l'entité phénoménologique. Cela n'empêche pas que l'on ait toujours de bonnes raisons de reconnaître, avec les empiristes britanniques, que *d'une certaine façon* le monde intérieur dépend des sources sensorielles.

La vision est la modalité sensorielle que les théoriciens considèrent presque toujours comme la source majeure de la connaissance perceptive, bien que nous ayons facilement recours au

toucher et à l'ouïe pour confirmer ce que nos yeux nous ont dit. Cette habitude que nous avons de voir tout dans l'esprit à travers la métaphore de la vision (une habitude à laquelle j'ai succombé deux fois dans cette phrase) conduit à bien des distorsions et des confusions, comme nous allons le voir. La vue domine tellement nos pratiques intellectuelles que nous avons de grandes difficultés à concevoir qu'il puisse en être autrement. Pour nous permettre de comprendre, nous faisons des diagrammes visibles et des tableaux. Pour « voir ce qui se passe » et « comment c'est possible », nous essayons d'imaginer dans notre « œil mental ». Une race de penseurs aveugles qui devrait se fier à son ouïe serait-elle capable de comprendre avec l'aide d'airs de musique, de tintements, et de gloussements perçus dans leur oreille mentale tout ce que nous saisissons grâce à des « images » mentales ?

Même les aveugles de naissance utilisent le vocabulaire visuel pour décrire leur processus de pensée. Pourtant, on ne voit pas clairement jusqu'à quel point cela résulte de leur penchant à utiliser les ressources d'expression du langage qu'ils tiennent de ceux qui ont la vue, ou d'une capacité à la métaphore qu'ils peuvent reconnaître en dépit des différences de leurs propres processus de pensée, ou même en faisant à peu près le même usage de la machinerie visuelle de leur cerveau que les gens dotés de la vue – en dépit du fait qu'ils n'ont pas les points d'entrée visuels appropriés. Si l'on pouvait répondre à ces questions, on aurait du même coup des lumières sur la nature de la conscience humaine normale, car elle se caractérise principalement par ce décor visuel.

Quand quelqu'un nous explique quelque chose, nous annonçons souvent ce que nous venons de saisir en disant « je vois ». Ce n'est pas là seulement une métaphore morte. La nature quasi visuelle de la *phénoménologie* de la compréhension a été presque entièrement ignorée par les chercheurs en sciences cognitives, en particulier en intelligence artificielle, qui ont essayé de créer des systèmes experts capables de comprendre le langage. Pourquoi ont-ils négligé la phénoménologie ? Probablement dans une large mesure parce qu'ils étaient convaincus que la phénoménologie, aussi réelle et aussi fascinante soit-elle, n'était pas fonctionnelle – et n'était qu'un rouage qui tourne mais qui n'enclenche aucune machinerie importante du processus de compréhension.

Les phénomènes dont font l'expérience différents auditeurs en réponse à la même énonciation peuvent varier quasiment à l'in-

fini sans différence apparente dans la compréhension ou dans le résultat. Prenez par exemple la variation dans l'imagerie mentale que pourrait provoquer chez deux personnes l'audition de la phrase :

Hier mon oncle a renvoyé son avocat.

Maurice pourrait commencer par se rappeler nettement les épreuves qu'il a subies *hier*, émaillées d'un coup d'œil rapide à un diagramme de la relation *être l'oncle de* (frère du père ou de la mère ; ou mari de la sœur du père ou de la mère), de l'image des marches d'un tribunal et de celle d'un vieillard en colère. Entre temps, peut-être, Mauricette sera passée sans forme d'image particulière sur « hier » et aura concentré son attention sur un aspect du visage de son oncle Albert, tout en imaginant une porte qui claque et le départ à peine « visible » de quelque dame bien mise nommée « avocate ». Tout à fait indépendamment des images mentales qu'ils ont pu produire, Maurice et Mauricette ont compris la phrase aussi bien l'un que l'autre. On peut s'en assurer par toute une batterie de paraphrases successives et de réponses à des questions. De plus, comme le feront remarquer ceux qui ont la fibre plus théorique, l'imagerie *ne pourrait pas* constituer la clef du processus de compréhension, parce que vous ne pouvez pas dessiner l'image d'un oncle, de la journée d'hier, d'un renvoi, d'un avocat. Les oncles, à la différence des clowns et des pompiers, ne diffèrent pas les uns des autres par des caractéristiques qu'on pourrait représenter visuellement, et les journées d'hier ne correspondent à rien de représentable. La compréhension, par conséquent, ne peut s'accomplir par un processus de conversion de toutes les données en un numéraire constitué d'images mentales, sauf à supposer que les objets représentés puissent s'identifier par des labels qui leur seraient attachés, auquel cas l'écriture de ces labels reposerait à nouveau sur des séquences verbales qu'il faudrait encore comprendre, ce qui nous ramènerait au point de départ.

Le fait que *j'entende* ce que vous dites dépend du fait que vous le disiez à une distance suffisante de mes oreilles à un moment où je suis éveillé, ce qui garantit que je l'entends. Le fait que *je comprenne* ce que vous dites dépend de nombreux facteurs ; mais cela ne dépend pas, semble-t-il, d'éléments identifiables de phénomènes internes ; aucune expérience consciente ne garantira que je vous ai compris ou que je vous ai mécompris. Le fait que

Mauricette se soit imaginé son oncle Albert ne l'empêche pas le moins du monde de comprendre que c'est l'oncle du locuteur, pas son propre oncle, qui a renvoyé son avocat ; elle *sait* ce qu'a voulu dire le locuteur ; elle se trouve seulement en train de se laisser aller à imaginer son oncle Albert, avec peu de risques de confusion, parce que sa compréhension de ce qu'a dit le locuteur ne dépend en rien de son imagerie ⁴.

On ne peut donc expliquer le phénomène de la compréhension en citant les traits d'une phénoménologie qu'il accompagnerait, mais cela ne veut pas dire que la phénoménologie soit absente. En particulier, cela ne veut pas dire qu'un modèle de la compréhension qui serait muet quant à la phénoménologie s'accorderait avec nos intuitions courantes sur la compréhension. Il ne fait pas de doute que le scepticisme que l'on manifeste couramment à propos de « la compréhension par des machines » des langues naturelles tire sa source du fait que de tels systèmes ne sont presque jamais dotés d'un espace de travail « visuel » dans lequel elles pourraient filtrer ou analyser les informations. Si elles en avaient un, on aurait bien plus le sentiment qu'elles comprennent effectivement l'information traitée (qu'il s'agisse ou non, comme certains le soutiennent, d'une illusion). Les choses étant ce qu'elles sont, si un ordinateur dit « je vois ce que vous voulez dire » en réponse à une information qu'on lui donne, nous sommes fortement tentés de rejeter cette assertion comme marque d'une fraude manifeste.

Cette tentation est certainement très forte. Par exemple, il est difficile d'imaginer comment qui que ce soit pourrait comprendre certains jeux de mots sans l'aide d'une imagerie mentale. Deux amis se tiennent au comptoir dans un bar, en train de boire ; l'un se tourne vers l'autre et dit : « Dédé, je pense que tu as dépassé la dose, ton visage devient tout embué ! »

N'avez-vous pas utilisé une image ou un diagramme momentané pour vous représenter l'erreur que le locuteur était en train de faire ? C'est là un exemple, semble-t-il, de *l'effet que cela fait de se mettre à comprendre quelque chose* : vous voilà, face à quelque chose de troublant ou d'indéchiffrable ou tout au moins d'inconnu – quelque chose qui d'une façon ou d'une autre crée une démangeaison épistémique, quand finalement vous dites : *Ça y est ! je comprends !* La compréhension s'éveille en vous, et la chose se trouve transformée ; elle devient utile, accessible, elle tombe sous votre contrôle. Avant le temps *t*, la chose n'était pas

comprise ; après le temps *t*, elle est comprise — c'est là un changement d'état clairement marqué et qu'on peut souvent dater précisément, même s'il s'agit, insistons bien là-dessus, d'une transition accessible subjectivement et découverte par introspection. C'est une erreur, comme nous allons le voir, d'en faire un modèle valable pour tout phénomène de compréhension ; mais il est certainement vrai que quand il y a un phénomène de début de compréhension (quand nous devenons conscients de commencer à comprendre quelque chose), c'est de ce phénomène qu'il s'agit.

L'idée d'imagerie mentale doit bien comporter quelque chose de correct, et si parler d'« images dans l'esprit » est une erreur, nous devons trouver la bonne façon de la caractériser. L'imagerie mentale affecte toutes les sortes de modalités sensorielles, pas seulement la vision. Imaginez « O douce nuit ! » en faisant bien attention de *ne pas* fredonner ou de ne pas chanter comme vous le faites d'habitude. N'est-il pas vrai cependant que vous « entendez » l'air dans votre esprit dans une clef particulière ? Si vous êtes comme moi, c'est ce qui se passe. Je n'ai pas la tonalité parfaite, et je ne peux donc vous dire quelle clef je viens d'imaginer, mais si quelqu'un devait maintenant se mettre à jouer « O douce nuit ! » au piano, je serais capable de dire avec certitude si c'était bien l'air que j'étais en train d'imaginer ou si ce que j'imaginai était trois tons au-dessus ⁵ ?

Non seulement nous nous parlons à nous-mêmes en silence, mais nous savons aussi parfois qu'il y a un « ton de voix » particulier. D'autres fois, il nous semble qu'il y a des mots, mais pas des mots *entendus* ; et d'autres fois, on ne trouve « là » que des ombres ou des esquisses de mots pour recouvrir nos pensées. Quand la psychologie introspective connaissait ses beaux jours, la discussion faisait rage sur la question de savoir s'il pouvait exister une pensée qui soit complètement « sans image ». Nous devons laisser ce problème de côté pour le moment, en nous contentant de remarquer que certaines personnes affirment avec certitude qu'il existe une telle pensée, alors que d'autres affirment avec tout autant de certitude qu'il n'y en a pas. Dans le chapitre suivant, nous proposerons une méthode permettant de résoudre ce genre de conflits. En tout cas, la phénoménologie de la pensée vive ne se limite pas au fait de *parler* à soi-même ; nous pouvons dessiner des images en imagination dans notre œil mental, conduire en imagination une auto à transmission manuelle, tou-

cher en imagination de la soie ou goûter en imagination un sandwich au beurre de cacahuète.

Que les empiristes britanniques aient eu raison ou pas de penser que ces sensations simplement imaginées (ou mémorisées) n'étaient que des copies affaiblies des sensations originales venues « de l'extérieur », elles peuvent produire en nous du plaisir et de la douleur tout comme les « vraies » sensations. Comme tout rêveur éveillé le sait, les fantasmes érotiques peuvent ne pas être un substitut entièrement satisfaisant de la chose réelle, mais ils sont néanmoins quelque chose qui nous manquerait beaucoup si l'on nous empêchait en quelque manière de les avoir. Non seulement ils peuvent produire du plaisir, mais ils peuvent aussi produire de véritables sensations et d'autres effets physiques bien connus. Nous pouvons pleurer en lisant un roman triste, tout comme les romanciers en en écrivant.

Nous sommes tous de bons connaisseurs des douleurs et des plaisirs que peut nous procurer l'imagination ; nous sommes nombreux à nous considérer nous-mêmes comme des experts dans la préparation de ces épisodes que nous apprécions tant ; mais nous serons sûrement surpris d'apprendre à quel point cette faculté peut devenir puissante quand elle fait l'objet d'un apprentissage sérieux. Je trouve stupéfiant, par exemple, qu'au cours de compétitions de composition musicale, les concurrents aillent souvent jusqu'à ne pas soumettre de bandes magnétiques ni d'enregistrements (ou d'exécutions publiques) de leurs œuvres ; ils soumettent des partitions écrites, et les juges portent avec certitude leurs jugements *esthétiques* sur la seule base de la lecture des partitions et de leur *écoute de la musique dans leur esprit*. Quelle est la valeur des meilleures imaginations musicales ? Est-ce qu'un musicien bien entraîné, qui lit aisément une partition, peut dire *juste* quel son peuvent produire des hautbois et des flûtes dissonantes par rapport à la masse sonore que produiront les cordes ? Les exemples ne manquent pas, mais pour autant que je sache, c'est là un territoire relativement inexploré, ouvert à des expérimentateurs habiles.

Les sensations imaginaires (si l'on peut appeler ainsi ces entités phénoménologiques) sont des objets tout à fait propices à l'appréciation et au jugement esthétiques. Mais alors pourquoi les sensations réelles comptent-elles nettement plus ? Pourquoi ne souhaite-on pas s'en tenir à des couchers de soleil en souvenir ou à des spaghetti au pistou seulement anticipés ? La majeure

partie du plaisir et de la douleur que nous associons aux événements qui nous arrivent sont, après tout, liés à nos anticipations et à nos souvenirs. Les simples moments de sensation ne sont qu'une partie de ce qui nous importe. La question de savoir pourquoi – et comment – les choses peuvent nous importer nous occupera dans les chapitres ultérieurs. Mais le fait que des sensations imaginées, ressouvenues, soient tout à fait différentes de sensations *affaiblies*, nous pouvons le souligner à travers une autre petite expérience de pensée, qui nous conduit au seuil de la troisième section du phénomène.

Affect

À présent, fermez les yeux et imaginez que quelqu'un vient de vous frapper très fort au tibia gauche (à peu près à trente centimètres au-dessus de votre pied) avec une botte à bout d'acier. Imaginez la douleur épouvantable avec autant de précision que possible ; imaginez qu'elle provoque des larmes sur votre visage ; imaginez que vous manquez de vous évanouir, tant la douleur que vous éprouvez est aiguë et puissante. Vous venez juste de l'imaginer très nettement. Avez-vous ressenti la moindre douleur ? Pourriez-vous simplement vous plaindre du fait que ce sont mes instructions qui ont provoqué en vous de la douleur ? Je m'aperçois que les gens répondent de façon très diverse à cet exercice, mais personne ne m'a jamais dit que l'exercice provoquait la moindre douleur réelle. Certains trouveront l'exercice plutôt troublant, alors que d'autres y verront un exercice mental plutôt amusant, bien moins désagréable que le plus doux des pincements au bras que vous appelleriez douleur.

Supposons maintenant que vous avez rêvé la même scène du coup de pied au tibia. Ce rêve peut vous causer un tel choc qu'il vous réveillera ; vous pourriez même vous retrouver en train de tenir votre tibia entre les mains et de geindre, les larmes aux yeux. Mais vous n'auriez ni inflammation ni marque de coup ou de contusion ; dès que vous seriez assez réveillé et d'aplomb pour vous fier à votre jugement, vous diriez qu'il n'y a pas trace de douleur dans votre tibia – s'il y en a jamais eu une. Les douleurs rêvées sont-elles des douleurs réelles, ou sont-elles des sortes de douleurs imaginées ? Ou quelque chose d'intermédiaire entre les deux ? Que penser des douleurs induites sous hypnose ?

Les douleurs rêvées et les douleurs induites par hypnose, tout au moins sont des états d'esprit qui nous importent réellement. Comparons-les, pourtant, aux états d'esprit que vous avez durant votre sommeil, quand vous vous retournez et pliez vos bras dans une position malencontreuse, et que, sans vous réveiller, sans en être pour le moins conscient, vous vous retournez pour prendre une meilleure position. S'agit-il de douleurs ? Si vous étiez éveillé, les états causés en vous par de telles contorsions seraient des douleurs. Il est des gens, heureusement peu nombreux, qui sont congénitalement insensibles à la douleur. Avant que vous commenciez à les jalouser, il vous faut savoir que du fait qu'elles ne font pas de corrections de posture pendant leur sommeil (ou quand elles sont éveillées !), elles deviennent très vite percluses de rhumatismes, leurs articulations étant endommagées parce qu'ils n'ont pas de sonnette d'alarme. Et ils se brûlent, se coupent et écourtent de diverses autres façons leur triste existence du fait de ce mauvais entretien de leurs organismes (Cohen et al., 1955 ; Kirman et al., 1968).

Il ne fait aucun doute que le système d'alarme constitué par les fibres productrices de la douleur est, avec les régions associées du cerveau, une aubaine que nous a procuré l'évolution, même si, en échange, certaines sonnettes d'alarme ne sont pas sous notre contrôle ⁶. Mais pourquoi les douleurs doivent-elles nous faire *si mal* ? Pourquoi n'y aurait-il pas une grosse cloche dans l'oreille, par exemple ?

Et à quoi peuvent bien servir la colère, la peur, la haine ? (Je pars du principe que l'utilité du désir pour l'évolution de l'espèce n'a pas besoin d'être défendue.) Pour prendre un exemple plus complexe, considérons la sympathie. Étymologiquement, le mot signifie *souffrir avec*. Les mots allemands correspondants sont *Mitleid* (douleur-avec) et *Mitgefühl* (éprouver-avec). Ou encore pensez à la *vibration sympathique*, dans laquelle une corde d'un instrument de musique se met à résonner du fait de la vibration d'une autre corde proche d'elle ou étroitement liée à elle, de sorte qu'elles partagent la même fréquence de résonance. Supposez que vous voyiez votre enfant dans un moment d'humiliation ou de dépression profonde ; c'est à peine si vous pouvez le supporter et des vagues d'émotion vous envahissent, noyant vos pensées, vous faisant perdre votre sang-froid. Vous êtes disposé à vous battre, à pleurer, à frapper. C'est un cas extrême de

sympathie. Pourquoi sommes-nous ainsi faits que de tels phénomènes se produisent en nous ? Et en quoi consistent-ils ?

Nous nous occuperons de la signification adaptative (s'il y en a une) de ces phénomènes dans plusieurs chapitres ultérieurs. Pour le moment, je veux seulement attirer l'attention, pendant notre petit tour d'horizon, sur l'importance indéniable de l'affect dans la conviction que nous avons que la conscience est quelque chose d'important. Considérons l'amusement, par exemple. Tous les animaux *désirent continuer à vivre*. Tout du moins s'efforcent-ils du mieux possible à se préserver eux-mêmes dans la plupart des conditions. Mais seul un petit nombre d'espèces nous paraissent vraiment capables de *goûter l'existence* ou de *s'amuser*. Nous avons à l'esprit des loutres fringantes qui glissent dans la neige, des lionceaux qui jouent, nos chiens et nos chats – mais pas les araignées ni les poissons. Les chevaux, tout du moins quand ils sont encore poulains, semblent être heureux de vivre, mais les vaches et les moutons ont l'air de s'ennuyer ou d'être indifférents. La pensée ne vous a-t-elle jamais traversé l'esprit que les oiseaux ne gagnent rien à voler, car peu d'entre eux semblent capables d'*apprécier* le caractère délicieux de leur activité ? L'amusement n'est pas un concept trivial, mais il n'a pas, à ma connaissance, assez attiré l'attention des philosophes. Nous n'aurons certainement pas d'explication complète de la conscience tant que nous n'aurons pas rendu compte du rôle qu'elle joue pour nous permettre de nous amuser (et nous seulement ?) Quelles questions faut-il poser ? Un autre exemple nous permettra de voir à quelles difficultés nous sommes confrontés.

Il existe une espèce de primate en Amérique du Sud, qui est plus sociable que la plupart des autres animaux et dont le comportement est curieux. Les membres de cette espèce se rassemblent souvent en groupes, petits ou grands, et durant les bavardages qu'ils ont entre eux, dans un grand nombre de circonstances, ils se trouvent amenés à produire des respirations involontaires, convulsives, une sorte de halètement bruyant, incontrôlé, renforcé mutuellement, et qui est quelquefois si violent qu'il les handicape. Loin d'être désagréables, cependant, ces attaques semblent être recherchées par la plupart des membres de l'espèce, et elles semblent même être une habitude très forte chez certains.

Nous pourrions être tentés de penser qu'à partir du moment où nous saurions quel effet cela fait d'être ces individus, en nous mettant à leur place, nous pourrions comprendre cette curieuse

addiction qu'ils éprouvent. Si nous pouvions « voir les choses de leur point de vue », nous saurions à quoi elle sert. Mais dans ce cas, nous pouvons être sûrs que les lumières que nous pourrions acquérir sur ce comportement ne permettraient pas de lever le mystère. Car nous avons déjà accès à ce que nous recherchons : l'espèce en question est l'*Homo sapiens* (il habite de fait en Amérique du Sud, entre autres lieux) et le comportement en question est le rire ⁷.

Aucun autre animal ne fait quoi que ce soit de ce genre. Un biologiste qui rencontrerait un phénomène aussi exceptionnel commencerait par se demander à quoi il sert. Ne trouvant pas d'analyse plausible mettant en évidence un avantage biologique direct lié au rire, il serait tenté d'interpréter ce comportement étrange et improductif comme le prix payé par l'organisme pour la possession d'un autre avantage. Mais lequel ? Que faisons-nous mieux grâce aux mécanismes sur lesquels repose notre penchant – et notre quasi-addiction – au rire et pour lesquels le prix vaut la peine d'être payé ? Le rire d'une certaine façon « nous évite-t-il le stress » qui s'accumule durant les efforts cognitifs compliqués qu'exige notre vie sociale avancée ? Mais pourquoi faudrait-il que ce soient des choses *amusantes* qui nous évitent le stress ? Pourquoi pas des choses *vertes* ou des choses *plates et simples* ? Ou encore, pourquoi ce comportement est-il le sous-produit de la propriété d'enlever le stress ? Pourquoi n'avons-nous pas le goût de rester debout en tremblant, en crachant, en nous grattant le dos les uns les autres, en fredonnant, en nous mouchant ou en nous léchant fiévreusement les mains ?

Notez bien que le *point de vue interne* est bien connu et ne pose pas problème. Nous rions *parce que nous sommes amusés*. Nous rions parce que des choses sont *drôles* – et le rire *sied* aux choses drôles à la différence du fait de se lécher les mains, par exemple. La raison pour laquelle nous rions est évidente (en fait *trop* évidente). Nous rions parce que nous éprouvons de la joie et du plaisir, parce que nous sommes heureux, parce que certaines choses sont hilarantes. S'il y eut jamais une *virtus dormitiva* dans une explication, c'est bien celle-là : nous rions à cause de l'hilarité provoquée par le stimulus ⁸. C'est certainement vrai ; il n'y a pas d'autre raison pour laquelle nous rions, quand nous rions sincèrement. L'hilarité est la cause constitutive du rire authentique. Tout comme la douleur est la cause constitutive du comportement

non feint de douleur. Comme c'est certainement vrai, nous ne devons pas le nier.

Mais il nous faut une explication du rire qui aille au-delà de cette vérité évidente, tout comme les explications habituelles de la douleur et du comportement de douleur vont au-delà de ce qui est évident. Nous pouvons donner une explication biologique tout à fait cohérente de la douleur et du comportement de douleur (en fait, nous venons juste de l'esquisser) ; nous voulons aussi une explication bien fondée de l'hilarité et du rire.

Et nous pouvons savoir d'avance que si nous parvenons effectivement à produire une explication de ce phénomène, elle ne satisfera pas tout le monde ! Certaines personnes, qui se considèrent comme des *antiréductionnistes*, se plaignent de ce que l'analyse biologique de la douleur *ne rende pas compte de ce qu'il y a de douloureux* dans celle-ci : elle laisserait de côté « le caractère horrible intrinsèque » de la douleur qui en fait ce qu'elle est. Ces gens critiqueront sans doute également toute explication du rire que nous pourrions trouver : elle laisse de côté ce qu'il a d'intrinsèquement hilarant. Ce genre de reproche est fréquent s'agissant de ce genre d'explications : « Tout ce que vous avez expliqué, ce sont le *comportement* et les mécanismes qui vont avec ; mais vous avez laissé de côté *la chose elle-même* : la douleur et ce qu'elle a d'horrible. » Cela pose des questions complexes, que l'on considérera à nouveau au chapitre 11, mais pour l'instant nous pouvons noter que toute analyse de la douleur qui *prendrait en compte* son caractère horrible serait circulaire – elle maintiendrait une *virtus dormitiva*. De même, une explication appropriée du rire *doit* laisser de côté l'hilarité intrinsèque présumée, le zeste, le caractère amusant de la chose, parce que leur présence n'aurait pour effet que de différer la tentative pour répondre à la question.

La phénoménologie du rire est hermétiquement scellée : nous voyons seulement directement, naturellement, sans faire d'inférence, avec une évidence qui va au-delà de l'« intuition », que le rire est ce qui va avec l'hilarité – que c'est la réaction « correcte » à l'humour. Nous pouvons parfois décomposer un petit peu le phénomène : la réaction correcte à quelque chose de drôle est l'amusement (un état interne de l'esprit) ; l'expression naturelle de l'amusement (quand il n'est pas important de la cacher ou de la réprimer, comme c'est quelquefois le cas) est le rire. Il semble que nous ayons à présent ce que les savants appellent

une variable intermédiaire, l'amusement, entre le stimulus et la réponse ; et il semble lié constitutivement aux deux bouts de la chaîne. C'est-à-dire que l'amusement est par-définition-ce-qui-est-provoqué-par-quelque-chose-de-drôle. Tout cela est évident. En soi, cela ne semble pas demander d'explication. Comme le disait Wittgenstein, les explications doivent s'arrêter quelque part. Mais tout ce que nous avons en réalité ici, c'est un fait brut – bien qu'explicable de façon claire – de la psychologie humaine. Nous devons aller au-delà de la pure phénoménologie si nous voulons expliquer l'un quelconque de ces habitants du jardin zoologique.

Ces exemples de phénoménologie, aussi divers soient-ils, semblent avoir deux caractéristiques importantes en commun. D'un côté, ils nous sont familiers : nous ne connaissons rien mieux que les entités de nos phénoménologies personnelles – tout au moins à ce qu'il semble. D'un autre côté, ils sont rétifs à une approche scientifique matérialiste : rien n'est moins semblable à un électron, à une molécule, à un neurone, qu'*un coucher de soleil pour moi maintenant* – du moins à ce qu'il semble. Ces deux traits ont à juste titre frappé les philosophes, et ceux-ci ont mis l'accent sur ce qu'ils ont de problématique de multiples manières. Aux yeux de certains, la grande énigme est cette familiarité particulière : comment pouvons-nous *ne jamais nous tromper* ou avoir un *accès privilégié* ou une *appréhension directe* de ces entités ? Quelle est la différence entre nos relations épistémiques avec notre phénoménologie et nos relations épistémiques avec des objets du monde extérieur ? Pour d'autres, le grand problème porte sur les étranges « qualités intrinsèques » – ou, pour employer le mot latin, les *qualia* – de notre phénoménologie : comment quelque chose qui est composé exclusivement de particules matérielles peut-il *être* l'amusement que je suis en train d'éprouver ? Comment cela peut-il avoir l'« homogénéité ultime » (Sellars, 1963) du morceau de glace rose que je suis en train d'imaginer ? Comment cela peut-il avoir pour moi l'importance que ma douleur a pour moi ?

Découvrir une explication matérialiste qui puisse rendre justice à tous ces traits ne sera pas facile. Mais nous avons fait quelques progrès. Notre bref inventaire a inclus quelques exemples dans lesquels une certaine connaissance des mécanismes sous-jacents menace – et peut-être usurpe – l'autorité que nous prêtons habi-

tuellement à l'introspection. En nous approchant un peu des vitrines, et en les regardant sous différents angles, nous avons commencé à rompre le charme, à dissiper la « magie » qui règne dans le jardin phénoménologique.

NOTES

1. Descartes, en fait, avait des animaux une conception semblable. Il soutenait que les animaux étaient en fait seulement des machines compliquées. Les corps humains, et même les cerveaux humains, n'étaient également pour lui que des machines. Seuls des esprits non mécaniques et non physiques pouvaient rendre des êtres humains (et des êtres humains seuls) intelligents et conscients. C'est là en fait une théorie subtile, que les zoologues d'aujourd'hui défendraient pour l'essentiel. Mais elle était trop révolutionnaire pour les contemporains de Descartes. Comme on peut l'imaginer, ils la caricaturèrent et la tournèrent ainsi en dérision. Des siècles plus tard, ces calomnies contre Descartes sont encore allègrement propagées par ceux qui considèrent comme inconcevable – ou tout au moins comme intolérable – le projet d'expliquer la conscience en termes mécanistes. Pour une analyse éclairante (cf. Leiber, 1988).

2. Cette question rhétorique implique, pour certains, la réponse tonitruante : Rien ! Par exemple, Mc Ginn (1989) défend cette réponse défaitiste en dressant une liste des options possibles qui ignore celles que nous envisagerons dans les chapitres suivants.

3. Pourquoi la *la* au-dessous du *do* m et le *la* au-dessus du *do* m (une octave plus haut) rendent-ils *le même son* ? Qu'est-ce qui fait que ce sont tous deux des *la* ? Quelle propriété ineffable de hauteur *la*-esque de son ont-ils en commun ? Quand deux tons sont distants d'une octave (et par conséquent sonnent « de la même manière » et ne sont différents que pour nous) la fréquence fondamentale de l'un est exactement le double de la fréquence fondamentale de l'autre. Le *la* standard en dessous du *do* m a 220 vibrations par seconde. Quand on les joue ensemble, des notes qui sont distantes d'une octave ou plus seront en phase. Cela explique-t-il le mystère de la parenté ineffable de ces sons ? « Pas du tout. Pourquoi les notes en phase de cette façon devraient-elles avoir *ce son* ? » Les notes qui ne sont pas en phase ne donnent pas *ce type de son*, mais elles rendent des sons semblables d'autres manières (par le timbre, par exemple), ce que l'on peut expliquer de diverses manières en termes de relations entre les fréquences des vibrations qu'elles produisent. Une fois que l'on a décrit les différentes manières par lesquelles les notes peuvent produire les mêmes sons ou des sons différents, et qu'on les a ordonnées selon leurs propriétés physiques et leurs effets sur notre système auditif, on peut même prédire avec une certaine précision, comment des notes nouvelles (par exemple des notes produites sur des synthétiseurs électriques) produiront pour nous certains sons. Si tout cela n'explique pas les parentés ineffables, que reste-t-il encore à expliquer ? (Je reviendrai sur ce thème bien connu au chapitre 11.)

4. On trouvera des développements classiques sur ce thème, avec divers autres arguments plus ou moins bons, dans les *Investigations philosophiques* de Wittgenstein (1953).

5. Un neurochirurgien m'a un jour parlé d'une opération qu'il avait pratiquée sur le cerveau d'un jeune homme atteint d'épilepsie. Comme c'est souvent le cas dans ce genre d'opération, le patient était totalement éveillé et il se trouvait seulement sous anesthésie locale, tandis que le chirurgien explorait délicatement la région exposée de son cortex, s'assurant que les parties qu'il essayait de retirer n'étaient pas absolument vitales en les stimulant électriquement et en demandant

au patient ce qu'il éprouvait. Certaines stimulations provoquaient des éclairs visuels ou amenaient le patient à lever sa main, d'autres provoquaient une sorte de sensation de bourdonnement, mais l'une des aires stimulées produisit une sensation délicieuse chez le patient : « C'est " Outta Get Me " de Gun's N' Roses, mon groupe de rock heavy metal favori. »

Je demandai au neurochirurgien s'il avait incité le patient à chanter ou à fredonner en même temps que la musique qu'il entendait, car il aurait été fascinant de savoir si le souvenir provoqué était en « haute fidélité » et jusqu'à quel point. Serait-ce la même clef et le même tempo que dans le disque ? Ce genre de chanson (à la différence de « Douce Nuit ») a une version canonique, en sorte que nous aurions pu simplement surimposer un enregistrement du fredonnement du patient avec le disque standard pour comparer les résultats. Malheureusement, en dépit du fait qu'un magnétophone ait fonctionné durant l'opération, le chirurgien n'avait pas demandé au patient de chanter en mesure. « Pourquoi ? », demandai-je. Il me répondit : « Je déteste la musique rock. »

Un peu plus tard dans la conversation, le neurochirurgien fit remarquer qu'il aurait à opérer à nouveau le même jeune homme ; j'exprimai l'espoir qu'il essaie de voir s'il pourrait à nouveau stimuler la musique rock, pour demander cette fois au type de chanter en même temps. « Je ne peux pas le faire, répondit le neurochirurgien, car j'ai coupé cette partie du cerveau. » « Était-ce une partie touchée par l'épilepsie ? », demandai-je. Il me répondit : « Non, je vous l'ai déjà dit, je déteste la musique rock ! »

La technique chirurgicale utilisée en la circonstance fut employée pour la première fois par Wilder Penfield il y a bien longtemps. Elle est décrite avec force détails dans *The Excitable Cortex in Conscious Man* (1958).

6. La littérature portant sur la justification évolutionniste de la douleur est criblée d'arguments étonnamment myopes. L'un des auteurs soutient qu'il ne peut pas y avoir d'explication évolutionniste de la douleur parce que certaines douleurs aiguës, comme la douleur provoquée par les calculs dans la vésicule biliaire, font résonner une alarme devant laquelle on était totalement impuissant jusqu'au développement de la médecine moderne. Aucun homme des cavernes n'a trouvé le moindre bénéfice reproductif dans la douleur qu'il éprouvait dans ses calculs. Par conséquent, la douleur – tout au moins certaines douleurs – serait un mystère dans l'évolution. Ce que cet auteur ignore, c'est le simple fait que pour avoir un système de douleur qui puisse vous avertir correctement de la présence de douleurs que vous voudriez éviter, comme celle d'une griffe ou d'un croc enfoncé dans votre ventre, vous obtiendrez très probablement le bonus – mais qui ne sera apprécié comme tel que plus tard – d'un système qui vous avertit également de la présence de crises que vous êtes incapables d'annuler. Et par la même occasion, il existe quantité d'états internes pour lesquels il serait aujourd'hui tout à fait désirable de bénéficier d'avertissements sous forme de douleur (le début d'un cancer, par exemple). Mais nous les oublions sans doute parce que notre passé évolutif ne les a pas associés à un avantage reproductif sous forme du câblage approprié (dût-il émerger par mutation).

7. « Que penserait un Martien rendant visite à un humain en train de rire ? Cela doit être vraiment affreux : la vue de gestes furieux, de membres flageolants, et du thorax animé de terribles contorsions » (Minsky, 1985, p. 280).

8. Dans la dernière pièce de Molière, la comédie classique du *Malade imaginaire* (1673), Argan, l'hypocondriaque, résout finalement ses problèmes en « devenant » un médecin pour pouvoir se soigner lui-même. Il n'a pas besoin de faire des études, mais seulement d'employer un peu de latin de cuisine. Dans un examen oral burlesque, il est mis à l'épreuve. Pourquoi, demande l'examineur, l'opium fait-il dormir les gens ? Parce que, répond le candidat, il a une *vertu dormitive* – le mot latin qui signifie « pouvoir de produire le sommeil ». « Bene, bene, bene, bene respondere », dit le chœur. Bien répondu ! Comme cela nous en apprend long ! Quelle perspicacité ! Et dans un esprit plus contemporain nous pourrions

seulement demander : qu'y a-t-il chez Catherine Deneuve qui la rend si bonne dans les films ? Elle est *photogénique*. Voilà la raison ! (Je me suis toujours demandé pourquoi.) On considérera plus en détail au chapitre 11 l'accusation de vacuité impliquée par le fait d'appeler une certaine explication une *virtus dormitiva*.

Chapitre 3

Une méthode pour la phénoménologie

La première personne du pluriel

On ne fait pas seulement de la zoologie sérieuse en se promenant à travers un zoo, en notant ceci et cela, et en admirant les curiosités. La zoologie sérieuse demande de la précision. Pour cela, il faut se mettre d'accord sur des méthodes de description et d'analyse, en sorte que d'autres zoologues puissent être sûrs qu'ils comprennent ce que vous êtes en train de dire. Pour être sérieuse, la phénoménologie a encore plus besoin d'une méthode claire et neutre de description, parce que, sans cela, deux personnes différentes n'utilisant, semble-t-il, pas les mots de la même manière, tout le monde serait un expert. Il est tout simplement stupéfiant de voir combien il est fréquent que les discussions « académiques » qui émaillent les controverses phénoménologiques dégénèrent en cacophonies et en coups de poings sur la table. Chacun parle sans écouter les autres. C'est d'autant plus surprenant, en un sens, que selon une tradition philosophique bien établie, *nous sommes tous d'accord* sur ce que nous trouvons quand nous « regardons à l'intérieur » de notre propre phénoménologie.

Faire de la phénoménologie a souvent semblé une pratique philosophique sans problème : elle consistait à regrouper des observations que tout le monde partage. Quand Descartes a écrit ses *Méditations* sur le mode du soliloque à la première personne, il s'attendait de toute évidence à ce que ses lecteurs s'accordent avec lui sur ses observations, en effectuant dans leur propre esprit les explorations qu'il décrivait et en obtenant les mêmes

résultats. Les empiristes britanniques Locke, Berkeley et Hume eux aussi écrivaient en présupposant que ce qu'ils faisaient, la plupart du temps, c'était *de l'introspection*, et que leurs lecteurs reproduiraient facilement leurs introspections. Locke a enrichi ce présupposé dans son *Essai sur l'entendement humain* (1690) en appelant sa méthode la « méthode historique simple ». Pour lui elle ne consistait pas en déductions abstruses ou en théories *a priori*, mais seulement en un relevé de faits observés, et Locke se contentait de rappeler à ses lecteurs ce qui était manifeste à quiconque regarderait au même endroit. En fait, à peu près tous les auteurs qui ont écrit sur la conscience ont pratiqué ce que nous pourrions appeler *la présomption de la première personne du pluriel*. Quels que soient les mystères que peut receler la conscience, *nous* (vous, gentil lecteur, et moi) pouvons parler confortablement ensemble de nos connaissances mutuelles, des choses que nous trouvons dans notre flux de conscience. À part quelques exceptions rebelles, les lecteurs ont toujours participé à la conspiration.

Tout cela ne poserait aucun problème si l'on ne se heurtait à un fait embarrassant : les thèses que l'on avance sous ces conditions d'accord mutuel et poli sont grasses de controverses et de contradictions. Nous nous méprenons tous à propos de quelque chose. Peut-être nous méprenons-nous quant au degré de ressemblance qui existe entre nous. Quand les gens rencontrent pour la première fois les différentes écoles de pensée qui s'affrontent au sujet de la phénoménologie, peut-être rejoignent-ils l'école qui leur paraît correcte. Peut-être chaque description propre à une école de phénoménologie est-elle fondamentalement correcte quant au type de vie intime qui est celle de ses membres, ce qui la conduit innocemment à généraliser à outrance, en faisant des analyses qui, par rapport à celles des autres, sont injustifiées.

Peut-être aussi nous méprenons-nous en prêtant à l'introspection et à nos pouvoirs personnels d'observation de notre propre esprit conscients la propriété d'être très fiables. Depuis Descartes et son « *cogito ergo sum* », cette capacité que nous avons nous a paru à l'abri de toute erreur ; nous avons un accès privilégié à nos propres pensées et sentiments. Et cet accès nous paraît meilleur que celui que peut avoir un observateur extérieur (« Imaginez quelqu'un essayant de vous dire que vous vous trompez sur ce que vous pensez et sentez ! »). Nous sommes « infailibles »

— nous sommes toujours sûrs d'avoir raison — ou tout du moins, nous sommes « impossibles à corriger » — que nous ayons raison ou tort, personne ne pourrait nous corriger (Rorty, 1970).

Mais peut-être cette doctrine de l'infailibilité n'est-elle qu'une erreur, aussi bien enracinée soit-elle. Peut-être, même si nous sommes fondamentalement semblables quant à notre phénoménologie, certains observateurs sont-ils tout simplement dans l'erreur quand ils essaient de la décrire. Mais puisqu'ils sont si certains d'avoir raison, il est relativement impossible de les corriger. (Ils sont impossibles à corriger en un sens dérogatoire.) D'une manière ou d'une autre, cela suscite des controverses. Il est encore une autre possibilité, qui à mon sens est plus proche de la vérité : nous nous abusons quand nous croyons que l'activité d'« introspection » consiste *toujours* à simplement « regarder et voir ». Je soupçonne que quand nous prétendons simplement utiliser nos pouvoirs d'*observation* intime, nous sommes en fait toujours en train de nous engager dans une forme de *théorisation* improvisée — et nous sommes des théoriciens incroyablement crédules, précisément parce qu'il y a si peu à « observer » et tellement de choses sur lesquelles on peut pontifier sans craindre d'être contredit. Quand nous introspectons, collectivement, nous sommes vraiment dans la même position que l'aveugle de la légende qui examinait différentes parties de l'éléphant. Cela semble de prime abord une idée absurde, mais essayons de voir ce que nous pouvons en faire.

Parmi les choses que nous avons rencontrées durant notre tournée d'inspection du phénomène au chapitre précédent, y en a-t-il une qui vous a surpris ? Avez-vous été surpris, par exemple, de ne pouvoir identifier la carte à jouer avant qu'elle soit placée juste au centre de votre champ de vision ? La plupart des gens, je l'ai découvert, sont surpris, même ceux qui savent que l'acuité de la vision périphérique est limitée. Si cela vous a surpris, cela veut dire que si vous aviez péroré sur ce sujet avant la démonstration surprenante, vous auriez eu de grandes chances de vous tromper. Les gens prétendent souvent avoir une connaissance intime d'une part plus grande du *contenu* de la périphérie de leur champ visuel qu'ils n'en ont en fait. Pourquoi s'aventurent-ils ainsi ? Pas parce qu'ils ont directement et infailiblement observé qu'ils ont des contenus périphériques de ce genre, mais parce que cela *leur paraît raisonnable*. Après tout, dans les conditions normales, vous ne relevez pas de lacune

blanche dans votre champ visuel. Et pourtant, s'il existait une région qui ne soit pas positivement colorée, vous noteriez la différence ; en outre, partout où vous regardez, vous voyez les choses en couleurs et dans leurs détails. Si vous pensez que votre champ visuel subjectif est fondamentalement une image intérieure composée de formes colorées, alors il est parfaitement raisonnable de supposer que chaque portion de la trame doit être colorée d'*une certaine couleur* — même une trame grossière comportant *une certaine couleur* ! Mais cette conclusion provient d'une conception douteuse de la nature de votre champ visuel ; ce n'est pas le produit d'une observation directe.

Je ne suis pas en train de dire que vous n'avez absolument pas d'accès privilégié à votre expérience consciente. Non, je suis en train de dire que nous tendons à penser que nous sommes bien plus à l'abri de l'erreur que nous ne le sommes en fait. Quand on leur montre les défauts de leur accès privilégié, les gens admettent en général qu'ils n'ont pas d'accès privilégié aux *causes* et aux *effets* de leurs expériences conscientes. Par exemple, ils peuvent être surpris d'apprendre qu'ils goûtent avec leur nez ou qu'ils entendent les notes basses avec leurs pieds ; mais ils n'ont jamais soutenu qu'ils avaient une autorité complète sur les causes ou les sources de leurs expériences. Ils n'ont d'autorité, disent-ils, que sur les expériences elles-mêmes, indépendamment de leurs causes et de leurs effets. Mais bien que les gens ne *disent* jamais qu'ils prétendent avoir une autorité seulement sur les contenus isolés de leurs expériences, et pas sur leurs causes et leurs effets, ils outrepassent souvent les limites qu'ils s'assignent. Par exemple, seriez-vous prêt à parier sur les propositions suivantes (j'en ai inventé au moins une) :

1. Vous pouvez avoir l'expérience d'une tache qui soit rouge et verte sur toute sa surface en même temps — une tache qui ait les *deux* couleurs (pas mélangées) en même temps.

2. Si vous regardez un cercle jaune sur un fond bleu, et si la luminosité ou le brillant du jaune et du bleu sont ajustés ensuite de manière à être égaux, la frontière entre le jaune et le bleu disparaît.

3. Il y a un son, qu'on appelle quelquefois le pôle auditif du barbier, qui semble continuer sans cesse à s'élever d'un ton, sans jamais s'élever plus haut.

4. Il existe une herbe qui, si vous en prenez une dose

excessive, vous rendra incapable de comprendre les phrases parlées de votre langue maternelle. Jusqu'à ce que l'effet disparaisse, votre audition est gênée, sans flou ni bruit supplémentaire, mais les mots que vous entendez vous paraissent appartenir à un langage entièrement étranger, bien que vous vous rendiez quelquefois compte que ce n'est pas le cas.

5. Si on vous masque les yeux et qu'un vibreur est appliqué sur votre bras pendant que vous touchez votre nez, vous sentirez votre nez s'allonger comme celui de Pinocchio ; si le vibreur est déplacé vers un autre point, vous aurez alors le sentiment bizarre de pousser votre nez à l'intérieur de vous – au point même que votre index semble finalement se tenir à l'intérieur de votre crâne.

En fait, j'ai inventé la proposition numéro 4. Mais, selon moi, elle est vraie. Après tout, dans la neuropathologie bien connue qu'on appelle la prosopagnosie, la vision du sujet est entièrement intacte et il lui est facile d'identifier la plupart des choses en les regardant, mais il ne parvient absolument pas à reconnaître les visages de ses amis et connaissances les plus proches¹. Pendant notre visite guidée du phénomène, je vous ai proposé un certain nombre d'expériences simples. Les phénoménologues tendent à soutenir que *parce que* nous n'avons pas d'autorité sur les causes et les effets physiologiques de notre phénoménologie, nous devrions les ignorer quand nous essayons de fournir une description pure, neutre, préthéorique de ce que nous découvrons comme étant le « donné » durant notre expérience quotidienne. C'est peut-être le cas, mais alors les habitants curieux du phénomène que nous ne rencontrerions jamais seraient nombreux ! Un zoologue qui chercherait à extrapoler toute la science de l'observation d'un chien, d'un chat, d'un cheval, d'un rouge-gorge et d'un poisson rouge, manquerait sans doute quelque peu son objectif.

La perspective à la troisième personne

Puisque nous allons nous autoriser à la pratique de la phénoménologie *impure*, il nous faut faire attention, plus que jamais, à la nature de notre méthode. Les phénoménologues adoptent en général la perspective à la *première personne* de Descartes, dans laquelle *je* décris, dans un monologue (que je vous laisse entendre)

ce que *je* trouve dans *mon* expérience consciente, en comptant sur *notre* accord. Mais j'ai essayé de montrer que le confort et la complicité qui résultent de la *perspective à la première personne* engendrent traîtreusement des erreurs. Dans l'histoire de la psychologie, en fait, la reconnaissance de plus en plus unanime de ce problème a conduit à la chute de l'introspectionnisme et à la naissance du behaviorisme. Les behavioristes ont pris garde minutieusement d'éviter toute spéculation sur ce qui se passe dans *mon*, dans *votre* ou dans *son* esprit. Ils ont défendu en fait la *perspective à la troisième personne*, selon laquelle seuls les faits engrangés « de l'extérieur » peuvent être pris comme des données. Vous pouvez faire un film vidéo des actions accomplies par des gens, puis mesurer les proportions d'erreurs dans les tâches impliquant un mouvement corporel ou les temps de réaction en poussant des boutons ou des leviers, le pouls, les ondes cérébrales, les mouvements oculaires, le rougissement (du moment que vous avez une machine qui mesure ces choses objectivement) et la réponse galvanique de la peau (la conductivité électrique détectée par les « détecteurs de mensonge »). Vous pouvez ouvrir la boîte crânienne des sujets (par la chirurgie ou par des techniques de scanner) pour voir ce qui se passe dans leur *cerveau*, mais vous ne devez pas formuler la moindre *hypothèse* quant à ce qui se passe dans leur *esprit*, car c'est là quelque chose sur quoi vous ne pouvez pas obtenir de données au moyen des méthodes vérifiables de manière intersubjective dans le cadre des sciences physiques.

En bref, puisque vous ne pouvez jamais « voir directement » à l'intérieur de l'esprit des gens, mais devez les croire sur parole ; les faits qui peuvent porter sur des événements mentaux ne font pas partie des données de la science, car on ne peut jamais vraiment les vérifier par des méthodes objectives. Ce scrupule *méthodologique*, qui est le principe de base de *toute* la psychologie expérimentale et des neurosciences d'aujourd'hui (et pas seulement de la recherche « behavioriste »), a trop souvent été élevé au statut de principe *idéologique*. Celui-ci a été formulé de la manière suivante :

Les événements mentaux n'existent pas (un point c'est tout ! C'est ce que l'on a souvent appelé « behaviorisme grossier »).

Les événements mentaux existent, mais ils n'ont aucun effet

causal, et par conséquent, ce ne sont pas des objets de science (épiphénoménisme – voir chapitre 11, section 5).

Les événements mentaux existent, et ont des effets, mais ces effets ne peuvent pas être étudiés par la science, qui devra se contenter de théoriser sur les effets et processus « périphériques » ou « inférieurs » du cerveau. (Cette conception est très répandue chez les neurologues, en particulier chez ceux qui se méfient des « théoriciens ». C'est en fait un dualisme ; ces chercheurs s'accordent apparemment avec Descartes sur le fait que l'esprit n'est pas le cerveau et se satisfont de n'avoir une théorie que pour le cerveau.)

Ces conceptions conduisent à des conclusions injustifiées. Même si les événements mentaux ne font pas partie des *données* de la science, cela ne veut pas dire que nous ne pouvons pas les étudier de manière scientifique. Les trous noirs et les gènes ne font pas partie des données de la science, mais nous avons de bonnes théories scientifiques à leur sujet. Le défi auquel nous sommes confrontés consiste à construire une théorie des événements mentaux en utilisant les données autorisées par la méthode scientifique.

Il faudra construire cette théorie à partir du point de vue de la troisième personne, parce que *toute* science relève de ce point de vue. Certains vous diront qu'une telle théorie de l'esprit conscient est impossible. On sait que le philosophe Thomas Nagel a soutenu les thèses suivantes :

« Il y a des choses au sujet du monde, de la vie et des humains qui ne peuvent être adéquatement comprises d'un point de vue optimalement objectif, quelle que soit la capacité qu'ait ce point de vue de nous faire aller au-delà de notre compréhension initiale. Une grande quantité de choses sont liées de manière essentielle à un certain point de vue, ou à un certain type de point de vue, et tenter de donner une analyse complète du point de vue en termes objectifs et détachés de ces perspectives conduit inévitablement à des réductions illusoire ou au refus d'admettre l'existence de certains phénomènes dont la réalité est évidente. » (Nagel, 1986, p. 7.)

Nous verrons. Il est prématuré de chercher à dire ce dont pourra ou non rendre compte une théorie tant que nous ne savons pas ce qu'elle dit effectivement. Mais si nous devons prêter une oreille équitable à une théorie, face à pareil scepticisme, nous aurons besoin de recourir à une façon neutre de *décrire les*

données – sans préjuger de la solution à ce problème. Il pourrait sembler que cette méthode n'existe pas, mais en fait il existe une méthode de ce genre. Je commencerai d'abord par la décrire, pour l'adopter ensuite.

La méthode de l'hétérophénoménologie ²

Voici un terme de mauvais augure : nous ne disons plus phénoménologie, mais *hétérophénoménologie*. De quoi peut-il bien s'agir ? C'est en fait quelque chose qui nous est familier à tous, mais il nous faut l'introduire avec une prudence excessive, en prenant note exactement de ce qu'elle présuppose et entraîne, parce qu'elle implique que l'on fasse un saut théorique énorme. Si on laisse de côté les raccourcis tentants, voici le chemin *neutre* qui conduit de la science physique objective et du privilège qu'elle accorde au point de vue de la troisième personne, à une description phénoménologique qui peut (en principe) rendre justice aux expériences les plus privées et les plus ineffables sans jamais abandonner les scrupules méthodologiques de la science.

Nous voulons une théorie de la conscience, mais on ne s'accorde pas sur la nature des entités qui sont douées de conscience. Les nouveau-nés humains en ont-ils une ? Et les grenouilles ? Que dire des huîtres, des fourmis, des plantes, des robots, des zombies, etc. ? Nous devrions rester neutres sur tous ces êtres pour le moment, mais il existe une classe d'entités dont tout le monde soutient qu'elle a des manifestations commentées. C'est celle de nos semblables, les humains adultes.

Or il se *peut* que certains de ces êtres humains adultes soient des zombies – au sens « technique » des philosophes. Le terme « zombie » vient apparemment du folklore haïtien et vaudou. Il fait référence, dans ce contexte, aux « morts vivants » punis pour des crimes et condamnés à errer, en grognant et en fixant le monde de leurs yeux morts, soumis inconsciemment aux prêtres vaudous ou aux chamans. Nous avons tous vu des zombies dans les films d'horreur, et on peut les distinguer immédiatement des gens normaux. (En gros, les zombies haïtiens ne peuvent pas danser, ni faire des jeux de mots, ni soutenir des discussions philosophiques animées, ni jouer leur rôle dans une conversation alerte – et ils sont tout simplement horribles ³.) Mais les philosophes utilisent le terme *zombie* pour désigner une catégorie

différente d'êtres humains imaginaires. Ils s'accordent pour dire qu'un zombie est ou serait un être humain qui manifesterait un comportement parfaitement naturel, alerte, loquace et vif, mais qui ne serait en fait pas conscient : c'est une sorte d'automate. Si les philosophes parlent de « zombies » en ce sens, c'est parce qu'on ne peut pas faire la distinction entre un zombie et une personne normale en examinant son comportement extérieur. Comme nous ne voyons jamais nos amis ni nos voisins sous un autre point de vue, *certaines de nos meilleurs amis pourraient être des zombies*. Tout au moins est-ce là la tradition, et je dois au départ ne pas me prononcer sur elle. Par conséquent, puisque la méthode que je décris ne forme pas d'hypothèse quant à la conscience effective d'un adulte humain quelconque – ou apparemment tel – elle va se concentrer sur cette classe d'adultes humains normaux, car si la conscience se trouve quelque part, c'est chez ces individus. Une fois que nous aurons déterminé les grandes lignes d'une théorie de la conscience humaine, nous pourrions nous tourner vers la conscience d'autres espèces, y compris les chimpanzés, les dauphins, les plantes, les zombies, les Martiens et les grille-pain automatiques (les philosophes s'autorisent souvent des fantaisies quand ils font des expériences de pensée).

Les êtres humains adultes (dorénavant, je dirai les gens) sont l'objet d'étude de nombreuses sciences. Les biologistes et les chercheurs en médecine, les nutritionnistes et les ingénieurs sondent leurs corps. (Ils se posent des questions du genre : à quelle vitesse les doigts humains sont-ils capables de taper à la machine ? Quelle est la force de tension d'un cheveu humain ?) Ce sont aussi les objets d'étude des psychologues et des neurologues, qui placent diverses personnes, appelées *sujets*, dans diverses situations expérimentales. Dans la plupart des expériences, les sujets doivent d'abord être rangés en catégories et préparés. On doit non seulement établir quel est leur âge, leur sexe, s'ils sont gauchers ou droitiers, quel est leur degré d'éducation, etc., mais on doit aussi *leur dire ce qu'ils doivent faire*. C'est là la différence la plus frappante entre les sujets humains et, par exemple, les cultures virales des biologistes, les échantillons de matériaux exotiques recueillis par les ingénieurs, les solutions du chimiste, les rats, les chats et les pigeons de l'éthologue.

Les gens sont les seuls objets d'étude scientifique dont la préparation requiert habituellement (mais pas toujours) la commu-

nication verbale. Cela relève en partie de l'éthique scientifique : on ne peut pas utiliser des gens dans des expériences sans leur consentement informé, et il est tout simplement impossible d'obtenir leur consentement informé sans avoir avec eux une interaction verbale. Plus important encore, de notre point de vue, est le fait qu'on utilise la communication verbale pour mettre en place les expériences et les conditionner. On demande aux sujets d'accomplir diverses tâches intellectuelles, de résoudre des problèmes, de rechercher des objets dans des dispositifs, de presser des boutons, de formuler des jugements, et ainsi de suite. La validité de la plupart des expériences dépend du fait qu'elles ont été préparées uniformément et correctement. S'il se révèle, par exemple, que les instructions ont été données en turc à des sujets dont la seule langue est le français, l'échec de l'expérience est tout à fait garanti. En fait, la mise en évidence ne serait-ce que de petits malentendus dans les instructions peut compromettre des expériences. C'est pourquoi il nous importe particulièrement que cette pratique qui consiste à préparer des sujets humains par la communication verbale soit validée.

Qu'est-ce que cette pratique verbale suppose ? C'est un élément indispensable aux expériences psychologiques. Mais présuppose-t-elle que les sujets sont conscients ? Les expérimentateurs ne se retrouvent-ils pas dans la même situation que les introspectionnistes ? Ne doivent-ils pas prendre au mot leurs sujets d'expérience, sans pouvoir vérifier ce qu'ils comprennent ? Ne courons-nous pas le risque de nous faire piéger par des zombies, des robots ou d'autres imposteurs ?

Pour le savoir, il est nécessaire d'examiner plus précisément les détails d'une expérience ordinaire sur un sujet humain. Supposez, comme c'est souvent le cas, qu'on fasse de nombreux enregistrements de l'ensemble de l'expérience : cassette audio et vidéo, électroencéphalogramme, et ainsi de suite. On ne comptera comme donné que ce qui est enregistré ainsi. Attachons-nous à l'enregistrement des sons – principalement des sons vocaux – qu'émettent les sujets et les expérimentateurs durant l'expérience. Comme les sons émis par les sujets sont produits par des moyens physiques, ils sont en principe explicables et prédictibles par la physique, conformément aux mêmes principes, lois et modèles que ceux que nous utilisons pour expliquer et prédire le bruit des automobiles ou les coups de tonnerre. Ou encore, puisque les sons sont produits physiologiquement, nous pourrions

utiliser les principes de la neurophysiologie pour tenter d'expliquer les sons, tout comme nous le faisons pour les crachotements, les ronflements, les gargouillements d'estomac et les craquements d'articulations. Mais il va de soi que les sons qui nous intéressent au premier chef sont les sons vocaux, et en particulier un sous-ensemble de ces sons (en ignorant les rots occasionnels, les reniflements et les bâillements) qui sont *en apparence* susceptibles d'une analyse linguistique ou sémantique. On ne sait pas toujours bien quels sons il faut inclure dans ce sous-ensemble, mais on peut s'en tirer sans encombre : on donne des copies des enregistrements sur cassettes à trois sténographes entraînés et on leur demande de préparer indépendamment les uns des autres des *transcriptions* des données élémentaires.

Cette simple étape est elle-même pleine de présupposés : elle nous fait passer d'un monde – celui de sons purement physiques – à un autre – celui des mots et des significations, de la syntaxe et de la sémantique. Cette étape conduit à une reconstruction radicale des données, abstraite de ses propriétés acoustiques et des autres propriétés physiques, et menant à des suites de mots (qui ont cependant encore des chronométrages précis – voir par exemple Ericsson et Simon, 1984). Qu'est-ce qui gouverne cette reconstruction ? Bien qu'on puisse présumer qu'il existe des relations régulières et détectables entre les propriétés de l'onde acoustique enregistrée sur la bande et les *phonèmes* que les dactylographes entendent et transcrivent sous forme de mots, nous n'en savons pas encore suffisamment sur ces relations pour les décrire en détail. (Si c'était le cas, le problème posé par la construction d'une machine à laquelle on pourrait dicter un texte serait résolu. Bien qu'on ait fait de grands progrès dans ce domaine, on se heurte encore à de grandes difficultés.) En attendant que cette recherche en acoustique et en phonologie soit plus complète, nous pouvons toujours faire confiance à nos transcriptions et les considérer comme des comptes rendus objectifs des données. Pour ce faire, quelques précautions élémentaires sont indispensables. En premier lieu, il faut que les sténographes préparent les transcriptions (plutôt que de se fier à l'expérimentateur pour faire ce travail, par exemple) : cela permet d'éviter des biais volontaires ou involontaires et des surinterprétations. (Les sténographes de tribunaux jouent le même rôle neutre.) Le fait de préparer trois transcriptions indépendantes nous donne une mesure de l'objectivité du processus. On peut penser que si l'enregistrement est

bon, les transcriptions s'accorderont mot pour mot sur tous les mots à un faible pourcentage près. Chaque fois que les transcriptions ne s'accordent pas, nous pouvons simplement rejeter les données si nous en avons envie, ou nous servir de deux ou trois transcriptions qui s'accordent pour fixer le résultat final.

La transcription ou le texte n'est pas, à strictement parler, *fournie* comme une donnée, car, comme nous l'avons vu, on la crée en soumettant les données brutes à un processus d'interprétation. Ce processus d'interprétation dépend d'hypothèses portant sur la nature du langage parlé en la circonstance, et sur certaines des intentions du locuteur. Pour mettre cela clairement en valeur, comparons la tâche que nous avons assignée aux sténographes à celle qui consiste à taper des transcriptions d'enregistrements de sons d'oiseaux ou de grognements de cochons. Quand le locuteur humain énonce « Vouzennui si j'pousse l'bouton 'vec la main gauch' », les sténographes s'accordent tous sur le fait qu'il a demandé : « Cela vous ennuie, si je pousse le bouton avec la main gauche ? » Mais c'est parce que nous connaissons le français, et parce que cet énoncé a un sens évident, dans le contexte. Et si le sujet *dit* « À présent, le point se meut de bauche à loite », nous autoriserons le sténographe à améliorer cela en tapant : « À présent, le point se meut de gauche à droite. » On ne dispose d'aucune stratégie semblable pour transcrire de manière purifiée les chants d'oiseaux ou les grognements de cochons – tout du moins tant qu'un chercheur n'aura pas découvert qu'il existe des normes régissant de tels bruits et n'aura pas codifié le système de description.

Nous « donnons un sens » sans effort – en fait involontairement – au flux sonore à travers le processus qui le transforme en mots. (Il vaut mieux autoriser les sténographes à changer « de bauche à loite » en « de gauche à droite », car ils le changeront sans doute sans même s'en rendre compte.) Le fait que le processus soit à la fois hautement fiable mais imperceptible dans des circonstances normales ne devrait pas nous masquer le fait que c'est un processus sophistiqué même quand il ne produit pas une compréhension mais s'arrête au niveau de la reconnaissance des mots. Quand le sténographe transcrit « À mes yeux, il y avait une sorte retentissante de cécité dans mon pressentiment, un courant sous-jacent révélateur d'anticipation de goût et d'affront, une multiplicité de confirmations anticipatoires qui ont révélé des surfaces derrière les surfaces », il peut ne pas avoir la moindre

idée de ce que cela peut vouloir dire. Mais il est parfaitement certain que c'étaient là les mots que le locuteur avait l'intention de prononcer et qu'il a réussi à énoncer, quoi qu'il ait pu vouloir dire.

Il est toujours possible que le locuteur n'ait eu aucune idée de ce que les mots voulaient dire. Le sujet, après tout, *pourrait* être un zombie, ou un perroquet habillé en humain, ou un ordinateur aux commandes d'un programme de synthèse de la parole. Ou, plus banalement, le sujet peut avoir été confus, ou sous l'influence d'une théorie mal comprise, ou en train d'essayer de jouer un tour à l'expérimentateur en émettant une série de non-sens. Pour le moment, selon moi, le processus de création d'une transcription ou d'un texte à partir de données est neutre par rapport à toutes ces possibilités étranges, même s'il part de l'hypothèse méthodologique qu'il existe un texte à déchiffrer. Quand on ne peut déchiffrer aucun texte, il vaut mieux rejeter les données portant sur ce sujet et repartir de zéro.

Pour le moment, la méthode décrite est nette, claire et incontestable. Nous avons atteint la conclusion mesurée suivante : nous pouvons transformer des enregistrements au magnétophone en textes sans abandonner la science. Nous avons pris notre temps pour nous assurer de ce résultat, parce que l'étape suivante, qui va nous permettre d'étudier empiriquement la conscience, est source de beaucoup d'obstacles et de confusions. Il nous faut aller au-delà du texte ; nous devons l'interpréter comme un enregistrement d'*actes de langage* ; il ne s'agit pas de simples prononciations ou de récitaions de mots, mais d'assertions, de questions, de réponses, de promesses, de commentaires, de demandes de clarification, de petites divagations à haute voix, de recommandations qu'on s'adresse à soi-même.

Cette forme d'interprétation nous conduit à adopter ce que j'appelle *la posture intentionnelle* ⁴ (Dennett, 1971, 1978a, 1987a) : nous devons traiter celui qui émet des bruits comme un agent, en fait comme un agent rationnel, qui arbore des croyances et d'autres états mentaux qui manifestent de l'*intentionnalité* ou de « l'être à propos de », et dont on peut expliquer (ou prédire) les actions sur la base du contenu de tels états. C'est pourquoi on doit interpréter les sons émis comme des choses que les sujets *avaient envie de dire*, comme des *propositions* qu'ils *avaient l'intention d'affirmer*, par exemple, pour diverses *raisons*. En fait, nous adoptions déjà ce genre d'hypothèses à l'étape précédente

quand nous purifions le texte. (Nous raisonnons ainsi : pourquoi quelqu'un *voudrait-il dire* « de bauche à loite » ?)

Quels que soient les dangers encourus si nous adoptons une posture intentionnelle vis-à-vis de ces comportements verbaux, ils sont le prix à payer si nous voulons avoir accès à quantité de truismes fiables que nous exploitons au moment de monter des expériences. De nombreuses raisons nous poussent à dire des choses, et il est important d'en exclure quelques-unes par un dispositif expérimental. Il arrive, par exemple, que les gens aient envie de dire certaines choses non parce qu'ils les croient, mais parce qu'ils croient que leur auditoire veut les écouter. D'ordinaire, il est important de commencer par diminuer la plausibilité de la présence de tels désirs : nous disons aux sujets que ce que nous voulons entendre est *tout ce qu'ils peuvent croire* et nous prenons soin de ne pas leur laisser savoir ce que nous espérons qu'ils croient. Autrement dit, nous faisons ce que nous pouvons pour les mettre dans une situation dans laquelle, étant donné les désirs que nous leur avons inculqués (le désir de coopérer, d'être payé, d'être un bon sujet), ils n'auront pas d'autre alternative que d'essayer de dire ce qu'ils croient.

Une autre application de la posture intentionnelle vis-à-vis de nos sujets est nécessaire si nous devons nous autoriser l'usage de types d'événements aussi utiles que peut l'être le fait de presser un bouton. Habituellement, presser un bouton est une manière d'accomplir un acte de langage conventionnellement déterminé, comme *affirmer* que les deux figures qu'on voit nous semblent surimposées juste *maintenant*, ou répondre que *oui*, mon jugement précipité, spontané (car on m'a dit que la vitesse de jugement faisait partie de l'essence de la tâche) est que le mot que je viens d'entendre était sur la liste que j'ai entendue il y a quelques instants. Par conséquent, les dispositifs expérimentaux exigent dans de nombreux cas que nous essayions de rendre explicites les significations de ces actions de pousser sur des boutons et de les incorporer dans un texte dont ils seront des éléments. La question de savoir quel est l'acte de langage que l'action de presser un bouton est supposée exécuter dépend de l'interprétation intentionnelle des interactions entre les sujets et l'expérimentateur qui faisaient partie de la préparation du sujet pour l'expérience. (Tous les actes de pousser un bouton ne constituent pas des actes de langage ; certains d'entre eux peuvent être

de faux déclenchements de tirs, ou de faux lancers de fusée, par exemple.)

Quand on doute que le sujet ait dit ce qu'il voulait dire, ou compris le problème, ou qu'il connaisse le sens des mots utilisés, nous pouvons demander des clarifications. On peut habituellement lever ces doutes. Dans l'idéal, ces mesures ont pour effet de dissiper toutes les sources probables d'ambiguïté et d'incertitude qui peuvent surgir dans la situation expérimentale, de façon à ce qu'une interprétation intentionnelle du texte (y compris les actions de presser le bouton) finisse par ne plus avoir de rivales crédibles. On la considère comme l'expression sincère, fiable, par un sujet *unique, unifié*, des croyances et opinions de ce même sujet ⁵. Comme nous le verrons, il existe cependant des cas dans lesquels cette hypothèse pose problème, en particulier quand nos sujets sont victimes d'une pathologie quelconque. Que faut-il faire, par exemple, des cas où les gens se plaignent d'être aveugles quand on a affaire à la soi-disant cécité hystérique, ou des cas où les sujets aveugles victimes d'anosognosie (dénî de cécité ou syndrome d'Anton) nient, de façon apparemment sincère, être aveugles ? Nous examinerons ces phénomènes dans les chapitres ultérieurs. Si nous voulons saisir le contenu des expériences de ces gens, il ne suffira pas de simplement les interviewer.

Les mondes fictionnels et les mondes hétérophénoménologiques

Outre les problèmes particuliers que posent les cas étranges, un problème général semble se présenter. La pratique d'interprétation du comportement verbal qui vient d'être décrite ne présuppose-t-elle pas elle-même la conscience du sujet ? Par conséquent, ne fait-elle pas une pétition de principe au sujet du problème des zombies ? Supposez que vous soyez confronté à un ordinateur « parlant » et que vous réussissiez à interpréter son « output » comme des actes de langage exprimant ses croyances et ses opinions, en supposant que celles-ci « portent sur » ses états conscients. Le fait qu'il *existe* une interprétation unique, et cohérente d'une suite de comportements n'implique pas que cette interprétation est *vraie* ; il se pourrait que les choses se passent seulement *comme si* le sujet était conscient ; nous risquons de

nous laisser abuser par un zombie dépourvu de toute vie intérieure. On ne pourrait pas *confirmer* que l'ordinateur était conscient de quoi que ce soit par cette méthode d'interprétation. C'est assez clair. Nous ne pouvons pas être sûrs que les actes de langage que nous observons expriment de véritables croyances quant à des expériences effectives ; elles n'expriment peut-être que des croyances *apparentes* au sujet d'expériences *non existantes*. Et pourtant, le fait de trouver une seule interprétation stable du comportement d'une chose quelconque comme acte de langage serait toujours un fait digne d'attention. Quiconque découvrirait une façon intersubjectivement uniforme d'interpréter le balancement des branches d'un arbre dans les branches comme des « commentaires » émis par le « temps » au sujet des événements politiques du moment aurait découvert quelque chose de merveilleux requérant une interprétation, même s'il se révélait qu'il s'est agi des effets d'un ingénieux dispositif créé par un ingénieur malicieux.

Heureusement, on peut disposer d'une analogie pour *décrire* ce genre de faits sans prétendre en même temps pouvoir les *expliquer* : nous pouvons comparer la tâche de l'hétérophénoménologue quand il interprète le comportement d'un sujet à celle d'un lecteur qui interprète une œuvre de fiction. Nous savons que certains textes, tels que des romans et des nouvelles, sont des fictions, mais cela ne nous empêche pas de les interpréter. En fait, à certains égards, cela rend la tâche d'interprétation plus facile : nous éliminons ou différons certaines questions difficiles portant sur la sincérité, la vérité ou la référence.

Considérons quelques faits incontestables quant à la sémantique de la fiction (Walton, 1973, 1978 ; Lewis, 1978 ; Howell, 1979). Un roman raconte une histoire, mais pas une histoire vraie, sauf par accident. Même si nous savons ou supposons que l'histoire racontée n'est pas vraie, nous pouvons parler de ce qui est *vrai dans l'histoire* ; c'est ce que nous faisons bien souvent. « Nous pouvons énoncer comme vrai que Sherlock Holmes vivait dans Baker Street et qu'il aimait à faire étalage de ses capacités intellectuelles. Nous ne pouvons pas énoncer comme vrai que c'était un père de famille dévoué, ou qu'il travaillait en collaboration étroite avec la police » (Lewis, 1978, p. 37). Ce qui est vrai dans l'histoire est bien plus large que ce qui est affirmé explicitement dans le texte. Il est vrai qu'il n'y a pas d'avions dans le Londres de Sherlock Holmes (bien que ce ne soit pas affirmé

explicitement ni même logiquement impliqué par le texte), mais il est également vrai qu'il existe des accordeurs de pianos (bien que – pour autant que je puisse m'en souvenir – aucun d'entre eux n'est mentionné, ni, là encore, sa présence impliquée logiquement par ce qui est dit). En plus de ce qui est vrai ou faux dans l'histoire, il existe une vaste zone indéterminée : il est vrai que Holmes et Watson ont pris le train de 11 heures 10 à la gare de Waterloo pour aller à Aldershot un jour d'été, mais il n'est ni vrai ni faux que ce jour était un mercredi (dans *L'Escroc*).

Des problèmes philosophiques délicieux se posent quant à la question de savoir comment dire (strictement) toutes les choses que nous voulons dire sans difficulté quand nous parlons de fiction. Nous ne nous en occuperons pas. Il est peut-être des gens que trouble profondément le statut métaphysique des personnages et des objets fictionnels, mais ce n'est pas mon cas. Mon optimisme béat me pousse à supposer que les réponses que nous devrions apporter, sur le plan ontologique, au sujet des résultats de la fiction ne posent pas le moindre problème philosophique profond : la fiction est de la *fiction* ; il n'y a pas de Sherlock Holmes. Mis à part les chicanes, donc, ainsi que les réponses techniques ingénieuses qu'on peut leur donner, je veux attirer l'attention sur un fait simple : on peut sans aucun doute interpréter la fiction, et cela produit certains résultats incontestables. En premier lieu, la mise en place de l'histoire, l'exploration du « monde de Sherlock Holmes », par exemple, n'est ni déplacée ni absurde ; on peut apprendre beaucoup sur un roman, sur son texte, sur ce qu'il entend montrer, sur l'auteur, et même sur le monde réel, en apprenant des choses sur *le monde décrit* par le roman. En second lieu, si nous faisons attention d'identifier et d'exclure des jugements de goût et de préférences (par exemple, « Watson est un poseur barbant »), nous pouvons amasser une grande quantité de faits objectifs incontestables au sujet du monde ainsi décrit. Tous les interprètes s'accordent pour dire que Holmes était plus malin que Watson ; l'objectivité tient à l'évidence sans mélange.

En troisième lieu – et c'est là un grand motif de soulagement pour les étudiants –, la connaissance du monde décrit dans un roman peut être indépendante de la connaissance du texte effectif du roman. Je pourrais probablement écrire un essai pour un examen sur *Madame Bovary*, mais je n'ai jamais lu le roman – même en traduction anglaise. J'ai vu la série télé sur la BBC,

donc je connais l'histoire. Je sais ce qui se passe dans ce monde. Cela montre, de manière générale que des faits qui portent sur le monde de la fiction sont des faits qui se situent à un *niveau purement sémantique* ; ils sont indépendants des faits syntaxiques du texte (si la fiction est un texte). Nous pouvons comparer la comédie musicale ou le film *West Side Story* avec la pièce de Shakespeare *Roméo et Juliette* ; en décrivant les différences et les similarités dans ce qui arrive dans ces mondes, nous voyons des similarités dans les œuvres d'art que l'on ne peut pas décrire dans des termes qui conviennent à la description syntaxique ou textuelle (sinon physique) des instanciations concrètes des fictions. Le fait que, dans chacun des mondes, il existe un couple d'amoureux qui appartiennent à différentes factions n'est pas un fait portant sur le vocabulaire, la structure des phrases, la longueur (en mots ou en bobine de film), ou la taille, la forme, ou une instantiation particulière physique de ces œuvres.

En général, on peut décrire ce qui est représenté dans une œuvre d'art (par exemple, *Madame Bovary*) indépendamment de la description de la *manière* dont la représentation est effectuée. (En général, bien sûr, on ne cherche pas à pratiquer cette séparation ; on mélange le commentaire au sujet du monde décrit avec un commentaire sur les moyens mis en œuvre par l'auteur pour mener à bien cette description, mais la séparation est possible.) On peut même imaginer que l'on en sache suffisamment sur le monde décrit pour être capable d'identifier l'auteur d'une fiction, quand on ignore le texte ou une quelconque traduction supposée fidèle. En apprenant indirectement ce qui se passe dans une fiction, on peut être prêt à dire : Seul Wodehouse aurait pu inventer une mésaventure aussi extravagante. Nous pensons pouvoir identifier des types d'événements et de circonstances (pas seulement des types de *descriptions* d'événements et de circonstances) comme kafkaïens, et nous sommes prêts à déclarer que certains personnages sont du pur Shakespeare. Beaucoup de ces convictions plausibles sont sans doute erronées (comme on pourrait le montrer en se livrant à d'ingénieuses expériences), mais ce n'est pas vrai de toutes. Je ne les mentionne que pour illustrer tout ce qu'on peut tirer de *ce qui est représenté*, même si on n'a qu'une connaissance très pauvre de *la façon dont la représentation est réalisée*.

Appliquons donc cette analogie au problème que rencontre l'expérimentateur qui veut interpréter les textes produits par des

sujets, sans préjuger si ses sujets sont ou non des zombies, des ordinateurs, ou s'ils mentent ou sont victimes de confusion. Considérez les avantages liés au fait d'adopter la tactique qui consiste à interpréter ces textes comme des sortes de fictions, non pas littéraires, mais comme des générateurs de *fictions théoriques* (lesquelles pourraient, bien entendu, se révéler vraies en définitive). Celui qui lit un roman laisse le texte *constituer* un monde (fictionnel), un monde déterminé par décret par le texte, entièrement extrapolé pour autant que faire se peut et indéterminé au-delà : notre expérimentateur, l'hétérophénoménologue, laisse le sujet du texte *constituer* le *monde hétérophénoménologique* de ce sujet, un monde déterminé par décret par le texte (en tant qu'interprété) et indéterminé au-delà. Cela permet à l'hétérophénoménologue de différer les problèmes épineux quant à ce que pourrait être la relation entre ce monde (fictionnel) et le monde réel. Cela permet aux théoriciens de s'accorder point par point sur *la nature* du monde hétérophénoménologique, tout en offrant des analyses totalement différentes des façons dont les mondes hétérophénoménologiques se projettent sur des événements dans le cerveau (ou l'âme, en l'occurrence). Le monde hétérophénoménologique du sujet sera un postulat théorique stable, intersubjectivement confirmable, ayant le même statut métaphysique que, par exemple, le Londres de Sherlock Holmes ou le monde selon Garp.

Comme dans la fiction, ce que dit l'auteur (l'auteur apparent) s'évanouit. Plus précisément, ce que dit l'auteur apparent fournit un texte qui, quand il est interprété selon les règles qui viennent d'être mentionnées, stipule ce qu'est un certain « monde ». Nous ne demandons pas comment Conan Doyle en est venu à connaître les couleurs du fauteuil de Sherlock Holmes et nous n'envisageons pas la possibilité qu'il ait pu se tromper ; nous corrigeons les erreurs typographiques et pour le reste nous mettons en place la lecture la meilleure et la plus cohérente du texte que nous puissions découvrir. De même, nous ne demandons pas comment les sujets (les sujets apparents) connaissent ce qu'ils affirment, et (à ce stade) nous ne supposons même pas qu'ils pourraient se tromper ; nous les prenons au mot (tel que nous l'interprétons). Notez également que, bien que les romans mentionnent souvent une restriction selon laquelle les descriptions qu'ils contiennent ne visent pas à décrire des gens réels, vivants ou morts, la tactique consistant à laisser un texte constituer le monde n'est pas limitée

aux œuvres littéraires *voulues* comme fictions par leurs auteurs ; nous pouvons décrire la reine Victoria d'un certain biographe, ou le monde d'Henry Kissinger, en négligeant totalement les intentions présumées des auteurs quant à la vérité de leurs descriptions ou quant au souci de faire référence, sans que ce soit fortuit, à des personnes existantes.

Le charme discret de l'anthropologue

Nous ne traitons pas ordinairement les gens comme des sources de fictions (théoriques). Le fait de simplement *concéder* une autorité constitutive à leurs déclarations a des allures un peu paternalistes, parce qu'on propose une sorte de pseudo-respect plutôt que du respect réel. On s'en aperçoit assez clairement quand on considère une application légèrement différente de la tactique des hétérophénoménologues, pratiquée par les anthropologues. On le verra bien sur un exemple. Supposez que les anthropologues découvrent une tribu qui croit en un dieu jusqu'à-présent-inconnu-de-la-forêt, appelé Phénhomme. En entendant parler de Phénhomme, les anthropologues sont confrontés à un choix fondamental : ils peuvent se convertir à la religion des indigènes et croire sincèrement à l'existence et aux bontés de Phénhomme, ou bien ils peuvent *étudier le culte* avec une attitude agnostique. Considérons la seconde voie, celle de l'agnosticisme. Sans croire en Phénhomme, les anthropologues décident néanmoins d'étudier et de systématiser du mieux qu'ils le peuvent la religion de ces individus. Ils établissent des descriptions de Phénhomme d'après leurs informateurs indigènes. Ils recherchent les points d'accord, mais ne les trouvent pas toujours (certains disent que Phénhomme a les yeux bleus, d'autres disent qu'il (ou elle) a les yeux bruns). Ils cherchent à expliquer et à éliminer ces désaccords, en identifiant et en ignorant ceux qui font leur malin, en explorant les reformulations possibles avec leurs informateurs, peut-être même en arbitrant les disputes. Petit à petit, une construction logique émerge : Phénhomme le dieu de la forêt, avec une liste complète de traits, d'habitudes, et une biographie. Les savants agnostiques (qui se nomment eux-mêmes des Phénhomménologues) ont décrit, ordonné, catalogué une partie du monde constituée par les croyances des indigènes ; s'ils ont fait leur travail d'interprétation correctement, ils ont compilé la descrip-

tion *définitive* de Phénhomme. Les croyances des croyants indigènes (nous pouvons les appeler les Phénhommanes) font autorité (c'est *leur* dieu, après tout), mais c'est seulement parce que Phénhomme est traité comme un « objet intentionnel », une simple fiction du point de vue des infidèles, et par conséquent entièrement comme une *créature* des croyances (vraies ou fausses) des Phénhommanes. Comme ces croyances peuvent être contradictoires, Phénhomme, en tant que construction logique, peut se voir attribuer des propriétés contradictoires – mais cela ne pose pas de problèmes aux yeux des Phénhomménologues parce qu'il n'est pour eux *qu'une* construction. Les Phénhomménologues essaient de présenter la meilleure construction possible, mais ils ne sont pas obligés à tout prix de résoudre toutes les contradictions. Ils sont prêts à découvrir des désaccords qui ne sont pas résolus et qui ne peuvent être éliminés au sein de la population des dévots.

Les Phénhommanes, bien sûr, ne voient pas les choses comme cela – par définition, puisque ce sont des croyants pour qui Phénhomme n'est pas un *simple* objet intentionnel, mais quelque'un d'aussi réel que vous ou moi. Leur attitude envers leur propre autorité au sujet des propriétés de Phénhomme est – ou devrait être – un peu plus complexe. D'un côté, ils croient bien qu'ils *savent* tout sur Phénhomme. Ce sont des Phénhommanes, après tout. Qui peut en savoir plus qu'eux ? Et pourtant, bien qu'ils se considèrent comme ayant une sorte d'infailibilité papale, ils admettent pouvoir se tromper sur certains points de détail. Ils pourraient ainsi peut-être apprendre quelque chose sur la nature véritable de Phénhomme. Par exemple, Phénhomme lui-même pourrait les corriger sur certains détails. Ils pourraient donc éprouver quelque gêne face à la crédulité suave (selon eux) des Phénhomménologues-chercheurs, qui les prennent presque toujours au mot, sans jamais mettre en question ni douter de quoi que ce soit, posant seulement des questions respectueuses en vue de résoudre des difficultés et des conflits apparents. Un indigène phénhommane qui perdrait la foi au contact des anthropologues et adopterait leur point de vue devrait adopter leur attitude de distance et de neutralité vis-à-vis de ses propres convictions (ne devrions-nous pas dire de ses propres convictions *antérieures*) et, dans l'affaire, quitter les rangs des vrais dévots.

La méthode hétérophénoménologique ne met en question ni n'accepte entièrement les assertions des sujets, mais elle main-

tient une neutralité constructive et sympathique, dans l'espoir de parvenir à une *description* complète du monde selon les sujets. Tout sujet qui serait mal à l'aise de se voir conférer pareille autorité constitutive pourrait protester : « Non, *vraiment* ! Ce que je vous décris est *parfaitement* réel ; cela a exactement les propriétés que j'affirme ! » En réponse, les Phénoménologues pourraient protester à leur tour en opinant du bonnet et en assurant au sujet que sa sincérité n'est pas mise en doute. Mais puisque les croyants veulent en général plus – ils veulent que leurs assertions soient crues, et à défaut, ils veulent savoir exactement quand leurs auditeurs cessent de les croire – il est en général de bonne politique pour les hétérophénoménologues, qu'ils soient anthropologues ou expérimentateurs étudiant la conscience en laboratoire, d'éviter d'attirer l'attention sur leur neutralité officielle.

La déviation par rapport aux relations interpersonnelles normales qui en résulte est le prix à payer pour qu'une science de la conscience soit neutre. Officiellement, nous devons garder l'esprit ouvert quant à la question de savoir si nos sujets apparents sont des menteurs, des zombies ou des perroquets habillés en personnes, mais nous n'avons pas à courir le risque de les troubler en claironnant que nous avons découvert le pot aux roses. Qui plus est, la tactique de neutralité n'est qu'une étape temporaire sur la voie de la construction et de la confirmation d'une théorie empirique qui pourrait en principe justifier les assertions des sujets.

À la découverte de ce dont il est vraiment question

Comment confirmer les déclarations des sujets quant à leur phénoménologie ? Nous verrons mieux quelles sont les possibilités grâce à nos analogies. Considérons la manière dont nous pourrions vérifier qu'un « roman » quelconque est en fait une authentique (ou largement authentique) biographie. Nous pourrions commencer par demander : de quelle personne bien connue du narrateur ce personnage s'inspire-t-il ? Ce personnage est-il la mère de l'auteur sous un certain déguisement ? Quels événements de l'enfance de l'auteur ont-ils été transférés dans cet épisode fictionnel ? Qu'est-ce que l'auteur est vraiment en train de dire ?

Demander à l'auteur pourrait bien ne pas être la meilleure manière de répondre à ces questions, car il se pourrait que celui-ci ne sache pas vraiment. Il est possible de soutenir dans certains cas que l'auteur – pour une raison quelconque – ne s'est pas permis de tirer un récit direct, non métaphorique des événements qu'il entendait raconter ; l'histoire qu'il a composée est un compromis ou un pur effet littéraire. Comme telle, elle se prête à une réinterprétation radicale (si nécessaire contre les protestations angoissées de l'auteur) de manière à révéler une histoire véridique, portant sur des personnes et des événements réels. Étant donné, pourrait-on soutenir, que ce n'est certainement pas une coïncidence si tel ou tel personnage a telles ou telles caractéristiques, nous pouvons réinterpréter le texte qui décrit le personnage de telle sorte que ses termes puissent apparaître comme faisant référence – de manière véritable, non fictionnelle – aux caractéristiques et aux actions d'une personne réelle. Le fait de décrire la Josiane de l'histoire comme une salope peut être considéré comme une diffamation envers la vraie Viviane, car tout ce qui est dit de Josiane s'applique *vraiment* à Viviane. Les protestations que pourra nous opposer l'auteur nous convaincront peut-être, à tort ou à raison, que la diffamation n'est pas, en aucune manière, une diffamation consciente ou délibérée, mais il y a longtemps que Freud et d'autres nous ont persuadés que les auteurs, comme n'importe qui, sont parfois aveugles aux ressorts profonds de leurs intentions. S'il peut exister des diffamations inconscientes, elles doivent pouvoir s'accompagner aussi de références inconscientes.

Ou encore, pour revenir à notre autre analogie, considérons ce qui se produirait si un anthropologue confirmait le fait qu'il existe réellement un individu aux yeux bleus nommé Phénhomme, qui guérit les malades et qui se balade de branche en branche dans la forêt, comme Tarzan. Cet individu n'est pas un dieu, et il n'est capable ni de voler ni de se trouver en deux endroits en même temps ; néanmoins, il est indubitablement la source effective de la plupart des témoignages, des légendes et des croyances des Phénhommanes. Cela entraînerait naturellement d'amères désillusions parmi les fidèles : certains réviseraient ou atténueraient leur credo ; d'autres s'en tiendraient à la version orthodoxe, même si cela impliquait l'association du « vrai » Phénhomme (gardant ses propriétés surnaturelles) avec son agent de chair et d'os vivant parmi nous. On peut comprendre

que les orthodoxes résistent à l'idée qu'ils aient pu se tromper à un tel point sur Phénhomme. À moins que le candidat des anthropologues destiné à jouer le rôle de référent réel de la doctrine phénhommaniste ait une ressemblance frappante, dans ses attributs et dans ses actes, avec le Phénhomme de la légende, ils n'auraient aucune raison de proposer une telle découverte. (Comparez avec la situation suivante : « J'ai découvert que le Père Noël existe réellement. C'est en fait un violoniste grand et mince qui vit à Miami sous le nom de Fred Dudley ; il hait les enfants et n'achète jamais de cadeaux à qui que ce soit. »)

Je suggère, par conséquent, que si nous devons trouver des événements réels dans le cerveau des gens qui aient *suffisamment* de propriétés « définitoires » caractérisant par ailleurs les entités qui peuplent les mondes hétérophénoménologiques, nous aurions le droit de penser que nous avons découvert ce dont ils parlent *réellement*. Même s'ils résistaient aux identifications et si nous découvriions que ces événements réels n'ont qu'une ressemblance très lointaine avec les entités hétérophénoménologiques, nous pourrions à bon droit déclarer que les gens se trompent tout simplement sur les croyances qu'ils expriment, en dépit de leur sincérité. Quelqu'un aurait toujours la possibilité d'insister, comme le Phénhomme endurci, et de dire que les vraies entités phénoménologiques accompagnent les événements sans pour autant s'identifier à eux. Mais c'est une autre question que de savoir si cette thèse est convaincante.

Comme les anthropologues, nous pouvons demeurer neutres tout en explorant ce domaine. Cette neutralité peut apparaître inutile. N'est-il pas tout simplement unimaginable que les savants puissent découvrir des phénomènes neurophysiologiques qui *s'identifient* purement et simplement aux entités célébrées par les sujets dans leurs hétérophénoménologies ? Les événements cérébraux semblent être trop différents des entités phénoménologiques pour être les vrais objets désignés par les croyances que nous exprimons dans nos comptes rendus d'introspection. (Comme nous l'avons vu dans le prélude, la substance mentale semble requise si elle doit être la matière dont les vaches violettes et autres entités phénoménales sont composées.) Je soupçonne que la plupart des gens continuent à trouver que la possibilité même de pratiquer cette identification est parfaitement imaginable. Mais plutôt que de concéder qu'elle est par là même impossible, je voudrais étendre les pouvoirs de notre imagination

encore un peu, en racontant une autre fable. Cette fable porte sur une entité phénoménologique particulièrement troublante, *l'image mentale* ; elle a l'avantage d'être une histoire vraie pour l'essentiel, bien que je m'autorise à la simplifier et à l'embellir un peu.

Les images mentales de Shakey

Au cours de la brève histoire des robots, Shakey, construit à l'Institut de recherche de Menlo Park à Stanford, en Californie, à la fin des années soixante par Nils Nilsson, Bertram Raphaël et leurs collègues, est légendaire. S'il le mérite, ce n'est pas parce qu'il a réussi quoi que ce soit ou parce qu'il a été capable de simuler de manière particulièrement réaliste un trait quelconque de la psychologie humaine, mais parce que à sa façon étrange, il a ouvert plusieurs possibilités intellectuelles et en a fermé certaines (Raphaël, 1976 ; Nilsson, 1984). C'était le type de robot qu'un philosophe pouvait admirer, une sorte d'argumentation à roulettes.

Shakey était une boîte montée sur roues avec un œil-télévision ; au lieu de promener avec lui son cerveau (à cette époque, c'était un gros ordinateur stationnaire), il lui était lié par radio. Shakey vivait dans des pièces, où les seuls objets à part lui étaient des boîtes, des pyramides, des plans inclinés et des plates-formes, soigneusement colorés et éclairés pour lui permettre de « voir » plus facilement. On pouvait communiquer avec Shakey en dactylographiant des messages sur un terminal attaché à son cerveau-ordinateur ; ils étaient écrits dans un vocabulaire sérieusement restreint de semi-anglais. Si on frappait « Enlève la boîte de la plate-forme », Shakey trouvait la boîte, localisait un plan incliné, poussait le plan incliné en position et le roulait jusqu'à la plateforme, pour finalement pousser la boîte.

Comment Shakey procédait-il ? Y avait-il (pourquoi pas ?) un nain humain à l'intérieur de Shakey, qui regardait l'écran de télévision et poussait les boutons de contrôle ? On aurait pu parvenir au résultat en trichant – en plaçant ainsi un homoncule unique et intelligent à l'intérieur. On aurait pu aussi charger un contrôleur humain en dehors de Shakey, de contrôler à distance les choses par radio. Ce serait la solution cartésienne, puisque le récepteur/transmetteur placé dans Shakey jouerait

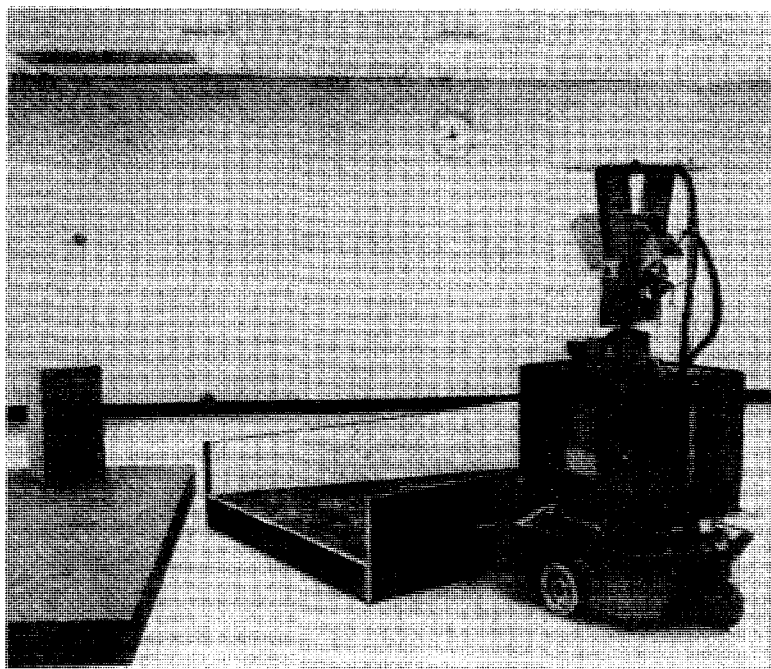


Figure 3.1

De *A Brief History of Time*,

© par Stephen W. Hawking, Bantam, Doubleday, Dell Publishing Group Inc.

alors le rôle de la glande pinéale et les signaux radio – fort peu miraculeux cette fois – tiendraient lieu des messages venus de l’âme dans la théorie de Descartes. Il est évident que ces « solutions » sont vaines. Mais existe-t-il une solution qui marche ? De prime abord, cela peut paraître inconcevable – ou tout au moins incroyablement complexe. Mais ce sont précisément ces obstacles d’imagination que nous devons affronter et vaincre. Il se révèle en fait plus facile que vous ne l’auriez supposé d’imaginer la manière dont Shakey pouvait accomplir ses exploits si l’on se passe d’un *homo ex machina*.

Comment, en particulier, Shakey faisait-il pour distinguer les boîtes des pyramides sans l’aide de son œil télévisé ? Les observateurs avaient déjà la réponse, en gros, quand ils pouvaient observer le processus sur un écran d’ordinateur. Une seule image de télévision, l’image d’une boîte par exemple, apparaissait sur le moniteur ; puis, elle était purifiée et affinée de diverses manières, jusqu’à ce que, merveilleusement, les limites de la boîte se des-

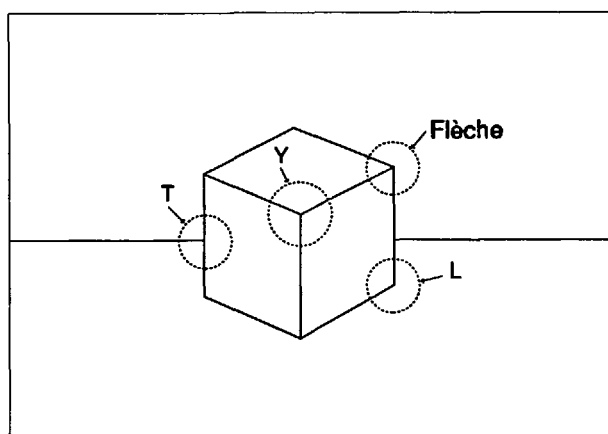


Figure 3.2

sinent en blanc – et que l'image tout entière devienne un dessin ligne à ligne (figure 3.3, p. 118).

Shakey se mettait alors à analyser le dessin ; chaque sommet était identifié comme un L, un T, un X, une flèche ou un Y. Si un sommet Y était découvert, l'objet ne pouvait être qu'une boîte, pas une pyramide ; d'aucun point de vue, une pyramide ne pouvait projeter un sommet en Y.

C'est une façon trop simplifiée de présenter les choses, mais elle illustre les principes généraux sur lesquels s'appuient les concepteurs du système : Shakey avait un programme de « sémantique de lignes » qui lui fournissait des règles générales permettant de déterminer la catégorie à laquelle appartenait l'objet dont l'image apparaissait sur le moniteur. En regardant le moniteur, les observateurs pouvaient s'attendre à éprouver un trouble soudain au moment de réaliser que quelque chose d'étrange était en train de se passer : ils étaient en train de regarder un processus de transformation d'images sur le moniteur ; mais *Shakey n'était pas en train de regarder ce processus*. De plus, Shakey ne regardait aucun autre moniteur sur lequel les mêmes images auraient été soumises à un processus de transformation et d'analyse. Il n'y avait pas de moniteur dans le matériel. En fait, le moniteur qu'ils étaient en train de regarder pouvait être éteint ou débranché sans que cela affecte les processus d'analyse perceptuelle de Shakey. Ce moniteur était-il une sorte de truc destiné à nous tromper ? Et au bénéfice de qui ? Seule-

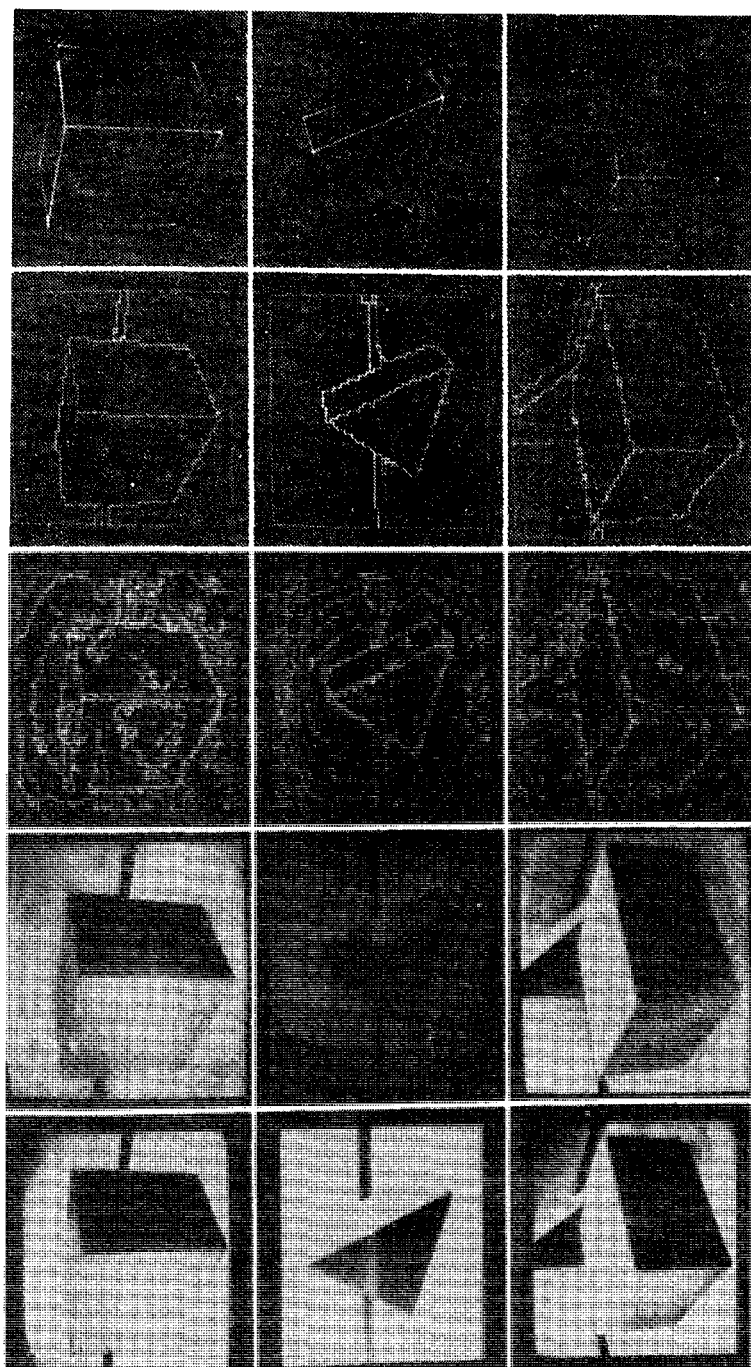


Figure 3.3

ment pour les observateurs. Quelle pouvait donc bien être la relation entre les événements que ceux-ci observaient sur le moniteur et les événements qui se produisaient à l'intérieur de Shakey ?

Le moniteur était destiné aux observateurs, mais l'idée même du moniteur était aussi destinée aux concepteurs de Shakey. Voyez combien leur tâche était pratiquement inimaginable : comment pourriez-vous prendre la sortie d'une simple caméra de télévision et en extraire en quelque sorte des identifications fiables de diverses boîtes ? De tous les kazillions de trames possibles que la caméra pouvait envoyer à l'ordinateur, seul un ensemble très réduit d'entre elles sont des images de boîtes ; chaque trame consiste en un tableau de cellules blanches et noires ou pixels, activées ou désactivées, correspondant à des zéros et à des uns. Comment écrire un programme qui identifierait toutes les trames qui seraient des images de boîtes et elles seules ? Supposez, pour simplifier à l'excès, que la rétine des caméras soit une grille de dix mille pixels, cent par cent. Dans ce cas, chaque image serait l'une des suites possibles de dix mille zéros et uns. Quelles structures dans les zéros et les uns pourraient bien s'accorder avec la présence de boîtes ?

Pour commencer, imaginez que l'on place tous ces zéros et ces uns sur un certain tableau, qui reproduise effectivement l'image de la caméra dans l'espace, comme dans les tableaux de pixels visibles sur le moniteur. Énumérez les pixels sur chaque rangée de gauche à droite, comme les zéros sur une page (à la différence des télévisions dans le commerce qui pratiquent des scannages en zigzag). Notez, ensuite, que les régions sombres sont principalement composées de zéros et que les régions claires sont composées de uns. De plus, une limite *verticale* entre une région claire à gauche et une région sombre à droite peut recevoir une description simple en termes de la suite de zéros et de uns : une suite de uns en majorité jusqu'au pixel numéro n , suivi d'une séquence d'une majorité de zéros, suivie d'exactly cent chiffres plus loin (dans la ligne suivante) par une autre suite composée principalement de uns jusqu'au pixel $n + 100$ suivie par une majorité de zéros, et ainsi de suite, par multiples de cent.

Un programme qui partirait à la chasse de ce genre de périodicités dans le flux des chiffres venant de la caméra serait capable de localiser des limites verticales de ce type. Une fois découverte, cette limite pourrait être transformée en une ligne blanche ver-

Les limites en pente sont juste un petit peu plus compliquées à mettre en place ; le programme doit chercher une progression dans la suite. Une fois que les frontières sont localisées et dessinées en blanc, le dessin est complet, et l'étape suivante, plus difficile, intervient : on « place » des « patrons » sur des petits bouts du segment de ligne de manière à pouvoir identifier les sommets. Une fois cette identification réalisée, l'usage du programme de sémantique des lignes pour catégoriser l'objet dans l'image n'est plus qu'une affaire de routine – dans certains cas, cela peut être une tâche aussi simple que de regarder un sommet en Y.

Cette démarche a plusieurs caractéristiques qui nous intéressent. En premier lieu, chaque sous-processus est « stupide » et mécanique. Cela veut dire qu'aucune partie du processus n'a besoin de comprendre ce qu'il fait ni pourquoi il le fait ; la réalisation mécanique de chacune des étapes ne comporte aucun mystère. Néanmoins, l'organisation habile de ces processus stupides, mécaniques, produit un dispositif qui *prend la place* d'un observateur capable d'enregistrer des informations. (Placez tout le système de vision dans une « boîte noire » dont la tâche serait de « dire à Shakey ce qu'il a besoin de savoir » au sujet de ce qu'il y a devant lui, à partir d'images télévisuelles comme entrées du processus. Au début, nous pourrions être enclins à penser que la seule manière d'y parvenir serait de placer un petit homme dans la boîte noire, qui regarde l'écran. Nous voyons à présent comment on peut remplacer cet homoncule, et son travail limité, par une machine.)

À partir du moment où nous voyons comment c'est possible, nous pouvons voir qu'alors que le processus est fortement *analogue* à celui qui consiste à regarder effectivement des points noirs et blancs sur l'écran (en dessinant et en gommant), la localisation effective dans l'ordinateur des opérations individuelles consistant à changer les zéros en uns et vice versa ne compte pas, du moment que les nombres qui sont des « adresses » temporaires des chiffres individuels codent l'information portant sur les pixels et leurs voisins. Supposez que nous éteignons le moniteur. Dans ce cas, bien qu'il n'y ait pas (ou pas nécessairement) d'image à deux dimensions localisable dans l'espace à l'intérieur de l'ordinateur (telle que, par exemple « trame d'excitation dans le matériel »), les opérations sont homomorphiques (parallèles) aux événements que nous sommes en train d'observer sur le moniteur. Ces événements sont authentique-

ment des images : une surface bidimensionnelle de points phosphorescents excités dessinant une forme avec une taille, une couleur et une orientation particulières. Par conséquent, en *un certain* sens précis, Shakey ne détecte pas de boîtes par une série de transformations d'images : la dernière image réelle dans le processus est celle qui apparaît sur le champ réceptif de la caméra. En un autre sens, strict mais néanmoins métaphorique, Shakey détecte bien des boîtes dans une série de transformations d'images – le processus qui vient d'être décrit qui transforme les frontières claires-sombres des objets en dessin d'une ligne puis classe les sommets. On peut comprendre pourquoi ce sens strict est néanmoins métaphorique si l'on remarque qu'il existe tout un ensemble de propriétés que possèdent les images *véritables* et que les images de Shakey n'ont pas : elles n'ont pas de couleurs, ni de taille, ni d'orientation. (Nous pourrions proposer une jolie petite énigme à propos de cette image : je suis en train de penser à une image qui n'est ni plus grande ni plus petite que Mona Lisa, qui n'est ni colorée ni en noir et blanc et qui n'est orientée dans aucune direction sur une boussole. De quoi s'agit-il ?)

Le processus que Shakey a utilisé pour extraire de l'information au sujet des objets à partir de la lumière de son environnement ressemble fort peu à celui de la vision humaine et probablement pas au processus visuel d'une créature quelconque. Mais nous pouvons ignorer ce fait pour le moment et envisager une possibilité assez abstraite quant à la manière dont les images mentales que rapportent les sujets peuvent être découvertes dans le cerveau. On a simplifié à l'excès l'analyse du système de vision de Shakey pour pouvoir dégager nettement les conclusions théoriques essentielles. Lançons-nous à présent dans un peu de science-fiction pour tirer une autre conclusion : supposons que nous fassions se rencontrer Shakey et un autre personnage fameux dans le domaine de l'intelligence artificielle, le SHRDLU de Terry Winograd (1972), qui manipulait des blocs (imaginaires) pour répondre ensuite à des questions au sujet de ce qu'il faisait et pourquoi. Les réponses de SHRDLU étaient pour l'essentiel « mises en boîte » ; c'étaient des phrases toutes prêtes déjà enregistrées et des canevas de phrases composées par Winograd. La raison d'être de SHRDLU était d'explorer de façon abstraite certaines tâches de manipulation d'information auxquelles serait confronté n'importe quel interlocuteur. Il ne s'agissait pas de modéliser de façon réaliste le comportement humain de discours, ce qui s'ac-

corde bien avec l'esprit de notre expérience de pensée. (Dans le chapitre 7 nous considérerons des modèles plus réalistes de production de discours.) Un échange avec notre nouvelle version de Shakey, reconstruit de manière à disposer d'un répertoire plus sophistiqué d'actions verbales, pourrait avoir l'allure suivante :

Pourquoi as-tu bougé le plan incliné ?

POUR POUVOIR ROULER SUR LA PLATE-FORME.

Et pourquoi voulais-tu faire cela ?

POUR POUSSER LA BOÎTE.

Et pourquoi voulais-tu faire cela ?

PARCE QUE TU M'AVAIS DIT DE LE FAIRE.

Mais supposez que nous demandions à présent à Shakey :

Comment distingues-tu les boîtes des pyramides ?

Que faudrait-il reconstruire dans Shakey pour qu'il nous « réponde » ? Voici trois possibilités :

1. Je passe en revue chaque suite de dix mille chiffres de long composée de zéros et de uns avec ma caméra, et je recherche certaines trames de suites, de manière à, etc. (la réponse sera très longue si nous laissons Shakey entrer dans les détails).

2. Je cherche les limites claires-sombres et je dessine des lignes blanches autour d'elles dans l'œil de mon esprit ; puis je regarde les sommets ; si je trouve un sommet en Y, par exemple, je sais que j'ai affaire à une boîte.

3. Je ne sais pas : certaines choses m'ont l'air d'avoir une allure de boîte. Cela me vient tout seul, par *intuition*.

Que doit dire Shakey ? Chaque réponse est vraie à sa façon ; ce sont des descriptions du traitement de l'information à différents niveaux de profondeur et selon différents grains. La question de savoir quelle est la réponse que nous pouvons dicter à Shakey dépend dans une large mesure de notre décision de lui donner telle ou telle capacité expressive d'accès (sa boîte noire SHRDLU) à ses processus perceptifs. Peut-être l'ingénieur aurait-il de bonnes raisons de dénier un accès profond (détaillé, dépendant du temps) aux processus intermédiaires d'analyse. Mais, quelles que soient les capacités d'autodescription que nous accordions à Shakey, la profondeur et la précision de la « connaissance » de ce qui se produit en lui qu'il pourra exprimer trou-

veront leurs limites. Si la meilleure réponse qu'il puisse fournir est (3), alors il est dans la même position quand il s'agit de dire comment il distingue les pyramides des boîtes que quand on nous demande comment nous distinguons le mot « soleil » du mot « sommeil » ; nous ne savons pas comment nous le faisons : l'un sonne pour nous comme « soleil » et l'autre comme « sommeil ». Nous ne pouvons faire mieux. Et si Shakey est conçu pour répondre (2), d'autres questions se posent encore auxquelles il ne pourra pas répondre, telles que : « Comment identifies-tu un sommet comme ayant la forme d'une flèche ? »

Supposez que nous construisions Shakey de sorte qu'il ait un accès de type (2) à ses processus perceptuels d'analyse ; quand nous lui demandons comment il s'y prend, il nous parle des transformations d'images auxquelles il se livre. À son insu, nous débranchons le moniteur. Sommes-nous autorisés alors à lui dire que nous avons une meilleure réponse ? Il n'est pas réellement en train de traiter des images, bien qu'il pense être en train de le faire ? (Il *dit* qu'il le fait, et de ce fait, en suivant la stratégie hétérophénoménologique, nous interprétons cela comme l'expression de sa croyance.) Si c'était une stimulation réaliste de ce qui se passe chez une personne, il pourrait bien rétorquer que nous n'étions pas en position de *lui* dire ce qui se passe dans son esprit ! Il savait ce qu'il était en train de faire, ce qu'il faisait *réellement* ! S'il était plus sophistiqué, il pourrait accorder que ce qu'il était en train de faire pourrait être décrit de manière allégorique comme du traitement d'images – bien qu'il se soit senti puissamment enclin à décrire ainsi ce qui se passe. Dans ce cas, nous serions capables de lui dire que sa façon métaphorique de présenter les choses était entièrement justifiée.

Si, d'un autre côté, nous étions plus diaboliques, nous pourrions nous arranger pour que Shakey parle de ce qu'il fait de façon totalement feinte. Nous pourrions faire en sorte qu'il veuille dire ce qui se passe en lui sans que cela ait la moindre relation à ce qui se passe effectivement. (« J'utilise les informations de mon récepteur télé pour guider un ciseau interne, qui taille une forme tridimensionnelle à partir d'un bloc d'argile mental. Puis, si mon homoncule peut s'asseoir dessus, c'est une boîte ; s'il tombe, c'est une pyramide. ») Ce compte rendu ne pourrait être interprété de manière véridique : Shakey serait seulement en train de *fabuler* – de raconter une histoire sans la « réaliser ».

Et cette possibilité, chez nous, montre pourquoi nous devons

nous donner la peine de passer par un détour : traiter l'hétérophénoménologie sur le même plan que l'interprétation de la fiction. Comme nous l'avons déjà vu, dans certaines circonstances, les gens se trompent tout simplement sur ce qu'ils sont en train de faire et sur la manière dont ils le font. Ce n'est pas qu'ils *mentent* dans la situation expérimentale, mais plutôt ils *fabulent* ; ils remplissent les vides, devinent, spéculent, confondent observer avec théoriser. La relation entre ce qu'ils disent et ce qui les conduit à dire ce qu'ils disent pourrait difficilement être plus obscure, à la fois pour nous autres, les hétérophénoménologues *qui voyons les choses de l'extérieur* et pour les sujets eux-mêmes. *Eux* n'ont pas le moyen de « voir » (avec un œil intérieur, on le présume) les processus qui gouvernent leurs assertions, mais cela ne les empêche pas d'avoir, du fond du cœur, des opinions à exprimer.

Pour nous résumer, les sujets, sans s'en rendre compte, créent des fictions, mais dire qu'ils ne s'en rendent pas compte, c'est admettre que ce qu'ils disent est, ou peut être, une analyse *de la façon dont les choses leur apparaissent exactement*. Ils nous disent *quel effet cela leur fait* de résoudre le problème, de prendre la décision, de reconnaître l'objet. Parce qu'ils sont sincères (en apparence), nous admettons que cela doit être l'effet que cela leur fait. Mais il s'ensuit que l'effet que cela leur fait est au mieux un guide incertain pour dire ce qui se passe réellement en eux. Il arrive que les fictions que nous créons sans nous en rendre compte se révèlent vraies en définitive, si nous admettons quelques ajustements métaphoriques comme nous l'avons fait avec la réponse de style (2) pour Shakey. Par exemple, les recherches récentes sur l'imagerie des psychologues cognitifs montrent que nos comptes rendus introspectifs sur les images mentales dont nous faisons l'expérience (qu'elles soient des images de vaches violettes ou pyramides) ne sont pas entièrement faux (Shepard et Cooper, 1982 ; Kosslyn, 1980 ; Kosslyn, Holtzman, Gazzaniga et Farrah, 1985). Ce point sera discuté plus en détail au chapitre 9, et nous verrons que nos comptes rendus introspectifs d'imagerie *peuvent* être interprétés comme véridiques. Mais, tout comme le Phénomène en chair et en os, qui se trouve en fait incapable de voler ou d'être à deux endroits en même temps, ce que nous découvrons effectivement dans le cerveau et que nous *identifions* comme des images mentales n'aura pas toutes les propriétés merveilleuses que les sujets attribuent en confiance

à leurs images. Les « images » de Shakey nous fournissent un exemple de la manière dont quelque chose qui n'était pas réellement une image pourrait être la chose même dont quelqu'un parlait comme d'une image. Les processus du cerveau qui sous-tendent l'imagerie humaine sont probablement très différents de ceux qui se déroulent à l'intérieur de Shakey. Mais nous avons ouvert un espace de possibilités qui était autrement difficile à imaginer.

La neutralité de l'hétérophénoménologie

Au début de ce chapitre, j'ai promis de décrire une méthode, la méthode hétérophénoménologique, neutre par rapport aux débats opposant les approches subjectives et objectives de la phénoménologie, et quant à la réalité physique ou non physique des entités phénoménologiques. Passons en revue cette méthode pour voir pourquoi c'est bien le cas.

Tout d'abord, qu'en est-il du problème des zombies ? Très simplement, l'hétérophénoménologie par elle-même ne peut pas distinguer les zombies des gens réels et conscients ; par conséquent, elle ne prétend pas résoudre le problème des zombies ou l'écarter. *Ex hypothesi*, les zombies se comportent exactement comme des gens normaux, et puisque l'hétérophénoménologie est une façon d'interpréter le comportement (y compris le comportement interne des cerveaux, etc.), elle parviendra exactement au même monde hétérophénoménologique pour Zoé et pour Zoé-zombie, son jumeau sans conscience. Les zombies ont un monde hétérophénoménologique, mais cela veut dire seulement que quand les théoriciens les interprètent, ils réussissent à effectuer exactement la même tâche, en utilisant exactement les mêmes moyens, comme quand nous interprétons les états mentaux de nos amis. Bien entendu, comme nous l'avons noté auparavant, certains de nos amis peuvent être des zombies. (Il m'est difficile de garder mon calme à ce sujet, mais puisque certains philosophes prennent le problème des zombies très au sérieux, je me sens obligé d'en faire autant.)

Il n'y a sûrement aucune difficulté, rien qui nous fasse sortir de notre neutralité, à admettre l'existence d'un monde hétérophénoménologique chez les zombies, puisque cela revient à admettre peu de chose. C'est une forme de minimalisme méta-

physique quant à l'hétérophénoménologie. La méthode décrit un monde, le monde hétérophénoménologique du sujet, dans lequel on trouve divers *objets* (des objets intentionnels, selon le jargon philosophique) et dans lequel diverses choses arrivent à ces objets. Si quelqu'un demande « Que *sont* donc ces objets, et *de quoi* sont-ils *faits* ? » La réponse *pourrait* être « Rien ! » De quoi est fait Mr Pickwick ? De rien. Mr Pickwick est un objet fictionnel, tout comme les objets décrits, nommés, mentionnés par l'hétérophénoménologie.

« Mais n'est-il pas un peu ennuyeux d'admettre, en tant que théoricien, que vous êtes en train de parler d'entités fictionnelles – de choses qui n'existent pas ? » Pas du tout. Les théoriciens de la littérature font un travail intellectuel honnête quand ils décrivent des entités fictionnelles, de même que les anthropologues qui étudient les dieux et les sorciers dans différentes cultures. C'est aussi bien entendu ce que font les physiciens qui, si on leur demande de quoi est fait un centre de gravité, diraient : « De rien ! Les objets hétérophénoménologiques sont, comme les centres de gravité ou l'Équateur, des *abstracta*, pas des *concreta* » (Dennett, 1987a, 1991a). Ce ne sont pas de vaines fantaisies, mais des fictions théoriques sérieuses. De plus, à la différence des centres de gravité, il reste toujours possible de les échanger au bénéfice de *concreta* si les progrès de la science empirique l'autorisent.

On peut étudier le Déluge de Noé de deux façons : on peut partir du principe que c'est un mythe pur et simple mais néanmoins hautement digne d'étude, ou bien on peut se demander si par-delà le mythe, il ne s'est pas effectivement produit une catastrophe météorologique ou géologique. Ces deux sortes d'analyses peuvent avoir un statut scientifique, mais la première est moins spéculative. Si vous choisissez la seconde, vous devez d'abord conduire une enquête serrée à partir de la première analyse pour déceler divers indices. De même, si vous voulez déterminer comment (ou même *si*) des entités phénoménologiques sont réellement des événements se produisant dans le cerveau, la première chose à faire consiste à dresser un catalogue hétérophénoménologique des objets. Cela risque d'offenser les sujets (de même, les anthropologues qui étudient Phénhomme risquent d'offenser leurs informateurs), mais c'est la seule manière d'éviter une bataille d'« intuitions » passant pour de la phénoménologie.

Pourtant, que faut-il répondre à l'objection selon laquelle l'hé-

térophénoménologie, du fait qu'elle part du point de vue de la troisième personne, laisse intact le *vrai* problème de la conscience ? Nagel, comme nous l'avons vu, insiste sur ce point, tout comme le philosophe John Searle, qui a lancé des recommandations explicites contre mon approche : « Rappelez-vous, admoneste-t-il, dans ces discussions, insistez toujours sur le point de vue à la première personne. La première étape du tour de passe opérationnaliste intervient quand nous essayons de nous figurer comment nous pourrions *savoir* quel effet cela fait d'être quelqu'un d'autre » (Searle, 1980, p. 451). Mais, ce n'est pas ce qui se passe. Notez que quand vous vous êtes placé entre les griffes hétérophénoménologiques, vous *gardez toujours le dernier mot*. Vous pouvez corriger, réviser et désavouer *ad libitum*. Tant que vous évitez de *théoriser* présomptueusement sur les causes ou sur le statut métaphysique des entités dont vous rendez compte, tout ce sur quoi vous insistez se voit conférer une autorité constitutive pour déterminer ce qui se passe dans votre monde hétérophénoménologique.

Si vous voulez que nous *croyions* tout ce que vous dites sur votre phénoménologie, vous ne demandez pas seulement à être pris au sérieux mais aussi à obtenir une infailibilité papale. C'est trop demander. Vous n'avez *pas* autorité sur ce qui se passe en vous, mais seulement sur ce qui *semble* se passer en vous, et nous vous donnons une autorité totale, dictatoriale sur l'analyse de la façon dont les choses vous apparaissent, sur *l'effet que cela vous fait d'être vous*. Et si vous déplorez que certaines des parties de ce qui vous apparaît soient ineffables, nous autres hétérophénoménologues, nous vous l'accorderons aussi. Quelles meilleures raisons pourrions-nous avoir de croire que vous êtes incapables de décrire quelque chose sinon que (1) vous ne le décrivez pas et que (2) vous avouez ne pas pouvoir le faire ? Bien entendu, vous pourriez être en train de mentir, mais nous vous donnerons le bénéfice du doute. Si vous rétorquez « Je ne suis pas seulement en train de dire que *je* ne peux pas le décrire ; je dis que c'est indescriptible ! », nous autres hétérophénoménologues, nous noterons qu'au moins, vous ne pouvez pas le décrire *maintenant*, et puisque vous êtes le seul en position de le décrire, c'est indescriptible en ce moment. Plus tard, peut-être, vous deviendrez capable de le décrire, mais bien sûr, à ce moment-là, *ce* sera quelque chose de différent, quelque chose de descriptible.

Quand je vous annonce que les objets de l'hétérophénomé-

nologie sont les fictions d'un théoricien, vous serez tentés (comme beaucoup, je crois) de vous précipiter sur cette affirmation et de dire :

« C'est là *seulement* ce qui distingue les objets de la vraie phénoménologie des objets de l'hétérophénoménologie. Mes objets *auto-phénoménologiques* ne sont pas des objets fictionnels — ils sont parfaitement réels, bien que je n'aie pas la moindre idée de ce dont ils sont faits. Quand je vous dis, sincèrement, que je suis en train d'imaginer une vache violette, je ne suis pas seulement en train de produire inconsciemment une suite de mots à cet effet (comme Shakey), habilement concoctée de manière à coïncider avec un événement physique vague, analogue dans mon cerveau ; je suis en train de rapporter consciemment et délibérément l'existence de quelque chose qui est *réellement là* ! Ce n'est pas une simple fiction théorique pour moi ! »

Prenez bien garde à ce discours. Vous n'êtes pas seulement en train de produire inconsciemment une suite de mots que vous énoncez. Vous êtes bien en train de produire une suite de mots ; vous n'avez pas la moindre idée de la manière dont vous le faites ou de ce qui rend possible cette émission. Mais, et vous insistez, vous ne faites pas *seulement* cela : vous savez *pourquoi* vous êtes en train de la faire ; vous *comprenez* la suite de mots, et vous *signifiez* ce que vous dites. D'accord. C'est pourquoi ce que vous dites marche si bien pour constituer un monde hétérophénoménologique. Si vous ne faisiez que répéter les mots plus ou moins au hasard comme un perroquet, les chances seraient infimes pour que cette séquence de mots produise une telle interprétation. Il existe certainement une explication satisfaisante de la raison pour laquelle vous dites ce que vous dites, une explication qui nous dit quelle est la différence entre simplement dire quelque chose et la dire en signifiant quelque chose, *mais vous n'avez pas encore cette explication*. Ou du moins pas toute. (Au chapitre 7, nous analyserons cette question.) Vous êtes probablement en train de parler de quelque chose de réel, du moins la plupart du temps. Voyons si nous pouvons trouver de quoi il s'agit.

Cela ne suffit pas à rassurer certaines personnes. Certains ne veulent pas jouer d'après ces règles. Certaines personnes dévotes et religieuses, par exemple, s'offensent quand leurs interlocuteurs ne font même que suggérer qu'il pourrait exister une autre religion vraie, qui soit différente de la leur. Pour ces personnes, l'agnosticisme n'est pas une attitude neutre ; c'est un affront,

parce que, d'après leur credo, être un mécréant est en soi un péché. Les gens qui croient de cette façon sont habilités à avoir leur croyance, et ils sont habilités (si c'est le mot juste) à avoir les sentiments blessés qu'ils éprouvent quand ils rencontrent des sceptiques ou des agnostiques, mais tant qu'ils ne peuvent pas maîtriser l'anxiété qu'ils ressentent quand ils apprennent que quelqu'un ne croit pas encore ce qu'ils disent, ils se soustraient à l'enquête scientifique.

Dans ce chapitre, nous avons développé une méthode *neutre* pour analyser et décrire la phénoménologie. Elle implique l'extraction et la purification des *textes* venant de *sujets* (apparemment) parlants, et l'utilisation de ces textes pour engendrer une fiction théorique, le *monde hétérophénoménologique* du sujet. Ce monde fictionnel est peuplé de toutes les images, événements, sons, odeurs, intuitions, pressentiments et sensations que le sujet croit (apparemment) sincèrement exister dans son flux de conscience. Si on l'étend au maximum, c'est un portrait neutre et exact de *l'effet que cela fait* d'être ce sujet – dans les termes mêmes du sujet, à partir de la meilleure interprétation possible en la circonstance.

Une fois qu'ils ont extrait cette phénoménologie, les théoriciens peuvent alors se tourner vers la question de savoir ce qui pourrait expliquer l'*existence* de cette hétérophénoménologie dans tous ses détails. L'hétérophénoménologie existe – c'est un fait aussi incontestable que l'existence des romans et autres fictions. Il ne fait pas de doute que les gens croient qu'ils ont des images mentales, des douleurs, des expériences perceptives, et ainsi de suite, et que *ces faits* – les faits qui portent sur ce que croient les gens, ce qu'ils rapportent quand ils expriment leurs croyances – sont des phénomènes que toute théorie scientifique de l'esprit doit analyser. Nous organisons nos données sur ces phénomènes en produisant des fictions théoriques : les « objets intentionnels » qui figurent dans des mondes hétérophénoménologiques. À partir de là, la question de savoir si les entités ainsi décrites existent en tant qu'objets, événements et états du cerveau – ou en l'occurrence, de l'âme – devient une question empirique. Si on découvre de bons candidats empiriques, nous pouvons les identifier comme étant les référents longtemps cherchés des termes employés par le sujet ; sinon, il nous faudra expliquer pourquoi il semble au sujet que ces entités existent.

À présent que nous avons mis en place nos principes, nous pouvons nous tourner vers la théorie empirique de la conscience

elle-même. Nous commencerons par envisager un problème portant sur la mesure et l'ordre temporel assigné aux entités qui figurent dans notre flux de conscience. Au chapitre 4, nous présenterons une esquisse de la théorie, et nous montrerons comment elle nous permet de réinterpréter certains phénomènes bien plus complexes qui ont suscité la perplexité des théoriciens. Les chapitres 6 à 8 étendront la théorie au-delà de cette esquisse initiale, en cherchant à dissiper les erreurs d'interprétation et les objections, et en illustrant d'autres vertus de la théorie proposée.

NOTES

1. Au sujet de la tache rouge et verte, voir Crane et Piantanida (1983) et Hardin (1988) ; au sujet de la limite disparaissante des couleurs, l'effet Lieberman (1927), voir Spillman et Werner (1990) ; pour le pôle du barbier auditif, voir Shepard (1964) ; pour l'effet Pinocchio, voir Lackner (1988). Pour plus de détails sur la prosagnosie, voir Damasio, Damasio et Van Hoesen (1982) ; Tranel et Damasio (1988) ; Tranel, Damasio et Damasio (1988).

2. Cette section et les suivantes reposent sur mes analyses antérieures des soubassements méthodologiques de l'hétérophénoménologie (Dennett, 1978c, 1982a).

3. Il y a quelques années, Wade Davis, un jeune anthropologue de Harvard, a annoncé qu'il avait déchiffré le mystère des zombies vaudous. Dans son livre, *The Serpent and the Rainbow* (1985), il a décrit la potion pharmacologique que préparaient les membres du culte vaudou de manière à donner aux humains des allures de cadavre ; après avoir été enterrés pendant plusieurs jours, ces infortunés individus sont parfois exhumés et on leur fait absorber un hallucinogène qui entraîne chez eux désorientation et amnésie. Le résultat de l'hallucinogène ou des lésions cérébrales causées par la privation d'oxygène pendant leur passage dans la tombe est qu'ils se mettent bien à errer à peu près à la façon des zombies dans les films, et à l'occasion ils peuvent être devenus des esclaves. En raison du caractère sensationnel des thèses de Davis (et du film qu'on a tiré de son livre romancé), ses découvertes ont suscité incrédulité et scepticisme dans certains milieux, mais ces critiques ont été écartées dans un second livre, plus sérieux sur le plan scientifique, *Passage of Darkness : The Ethnobiology of the Haitian Zombie* (1988). Voir aussi Booth (1988) et Davis (1988b).

4. *Intentional stance*. Dans *La Stratégie de l'interprète*, j'ai traduit par « perspective intentionnelle » ou « point de vue intentionnel », et non comme ici par « posture intentionnelle », en raison notamment des contextes où « posture » a un sens trop net de « posture physique » (y compris quand il est question de la *physical stance* : « posture physique » eût été ambigu). Je traduis de même ici par « point de vue » dans certains cas (NdT).

5. Dans « How to change your Mind » (*Brainstorms* 1978a), j'adopte un usage conventionnel du mot « opinion », qui me permet de faire une distinction entre les croyances proprement dites et d'autres états plus linguistiques, que j'appelle des opinions. Les animaux sans langage peuvent avoir des croyances, mais pas des opinions. Les humains ont les deux, mais si vous croyez que demain on sera vendredi, ce sera selon ma terminologie votre opinion que demain sera vendredi. On ne peut pas avoir ce type d'état mental sans langage. Bien que je ne présuppose pas ici une connaissance de cette distinction, j'entends que mes affirmations s'appliquent à ces deux catégories.

DEUXIÈME PARTIE

Une théorie empirique
de l'esprit

Chapitre 4

Versions multiples contre Théâtre Cartésien

Le point de vue de l'observateur

« Il n'y a pas de cellule ou de groupe de cellules dans le cerveau qui ait une prééminence anatomique ou fonctionnelle telle qu'elle puisse apparaître comme la pierre de fondation ou le centre de gravité de l'ensemble du système. »

William James, 1890.

En général, les plaisanciers qui naviguent le long d'une côte dangereuse s'assurent qu'ils sont en sécurité en cherchant un repère. Ils choisissent un récif visible au loin et situé dans la direction où ils veulent se rendre ; ils vérifient leur carte pour voir s'il n'y a pas d'obstacles cachés sur la ligne droite entre le repère et l'endroit où ils se trouvent, et ils mettent directement le cap vers lui. Pendant une heure ou deux, le skipper cherche à se diriger directement vers le repère, en corrigeant les erreurs. Mais il arrive souvent que les skippers s'obnubilent tellement sur ce projet qu'ils oublient de virer à la dernière minute et vont droit sur le récif ! Certains de parvenir à leur objectif plus limité – le repère –, ils en oublient leur objectif général – échapper au danger. Dans ce chapitre nous allons voir comment certains des paradoxes les plus troublants de la conscience naissent parce que nous nous cramponnons trop longtemps à une habitude de pensée qui nous permet d'éviter les écueils.

Qui dit esprit conscient dit *point de vue*. C'est l'une des idées

fondamentales quand il est question de l'esprit – ou de la conscience. Un esprit conscient est un observateur, qui ne prend en compte que certaines des informations qui sont disponibles. Il prend en compte l'information qui est disponible pendant une suite (en gros) continue de temps et de lieux dans l'univers. Pour faciliter les choses, nous allons considérer que le point de vue d'un sujet conscient particulier est justement cela : un *point* qui se meut à travers l'espace-temps. Considérons, par exemple, les diagrammes usuels en physique et en cosmologie, qui illustrent l'effet Doppler ou les effets de torsion de la lumière dus à la gravitation.

L'observateur, dans la figure 1, est fixé en un point sur la surface de la terre. Pour des observateurs situés à des points différents de l'univers, les choses se présenteraient de façon différente. Nous expliquons l'étonnante différence temporelle entre le son de différents feux d'artifice et le moment où nous les voyons en relevant les différentes vitesses de transmission du son et de la lumière. Elles arrivent à *l'observateur* (au point qu'il occupe) à différents moments, même si elles ont quitté la source au même moment.

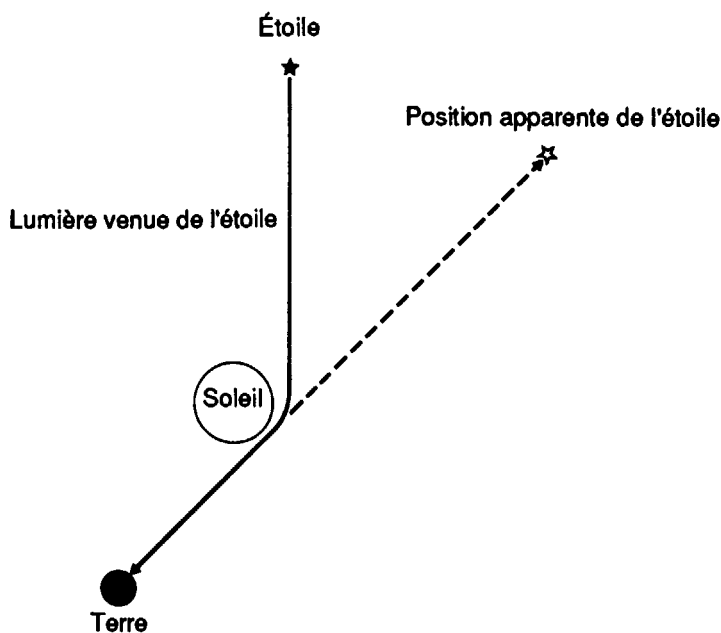


Figure 4.1

Mais que se passe-t-il quand nous nous concentrons sur l'observateur et essayons de localiser plus précisément le point de vue de l'observateur, comme un point situé à l'intérieur de l'individu ? Les hypothèses simples qui marchent si bien à plus grande échelle semblent ne plus opérer ¹. Il n'existe pas de point unique dans le cerveau par lequel toute l'information arrive. Les conséquences de ce fait sont loin d'être évidentes : elles vont contre nos intuitions.

Comme nous allons considérer des événements qui se produisent à une échelle relativement microscopique de l'espace et du temps, il est important d'avoir une idée claire des grandeurs impliquées. Toutes les expériences que nous allons envisager mettent en jeu des intervalles de temps mesurés en millisecondes ou en millièmes de secondes. Que représentent 100 ou 50 millisecondes ? Vous pouvez énoncer environ quatre ou cinq syllabes par seconde ; de même, un film envoie une image toutes les 42 millisecondes (en fait toute image est stationnaire et exposée *trois* fois pendant 42 msec, pour des durées de 8,5 msec avec 5,4 msec de noir entre chacune d'elles). Aux États-Unis, la télévision émet trente images par seconde, ou une image toutes les 33 msec. (En fait, chaque image est envoyée deux fois et empiète sur celle qui la précède.) En bougeant votre pouce aussi vite que possible, vous pouvez faire démarrer et éteindre une montre électronique en environ 175 msec. Quand vous vous tapez sur le doigt avec un marteau, les fibres nerveuses rapides (activées à la myéline) mettent environ 20 msec pour envoyer un message au cerveau ; les fibres C lentes, sans myéline, envoient des signaux de douleur qui mettent bien plus de temps – 500 msec – pour couvrir la même distance.

Voici un tableau des valeurs approximatives en millisecondes de certaines durées importantes :

Dire « Un, Mississipi »	1 000 msec
Fibre sans myéline, du petit doigt au cerveau	500 msec
Une balle de pénalty à 120 km arrive sur le goal	330 msec
Énoncer une syllabe	200 msec
Mettre en marche et arrêter un chronomètre	175 msec
Une image dans un film	42 msec
Une image télévision	33 msec
Une fibre rapide (avec myéline), du doigt au cerveau	20 msec
Le cycle de base d'un neurone	10 msec
Le cycle de base d'un ordinateur personnel	0,0001 msec

Descartes, l'un des premiers à avoir réfléchi sérieusement sur ce qui doit se produire quand nous regardons de près à l'intérieur du corps de l'observateur, élaborait une idée qui apparaît de prime abord si naturelle et si séduisante qu'elle continue encore à influencer sur notre conception de la conscience. Comme nous l'avons vu au chapitre 1, pour Descartes, le cerveau *avait bien* un centre : la glande pinéale, qui servait de porte d'entrée à l'esprit conscient (voir figure 1.1, page 51). La glande pinéale est le seul organe du cerveau qui ne soit pas apparié à un des deux hémisphères, mais qui soit situé sur la ligne médiane. Elle est désignée par « L » dans le diagramme qui suit, dessiné par le grand anatomiste du xvi^e siècle Vésale. Plus menue qu'un petit pois, elle se tient, dans un splendide isolement sur son socle, reliée au reste du

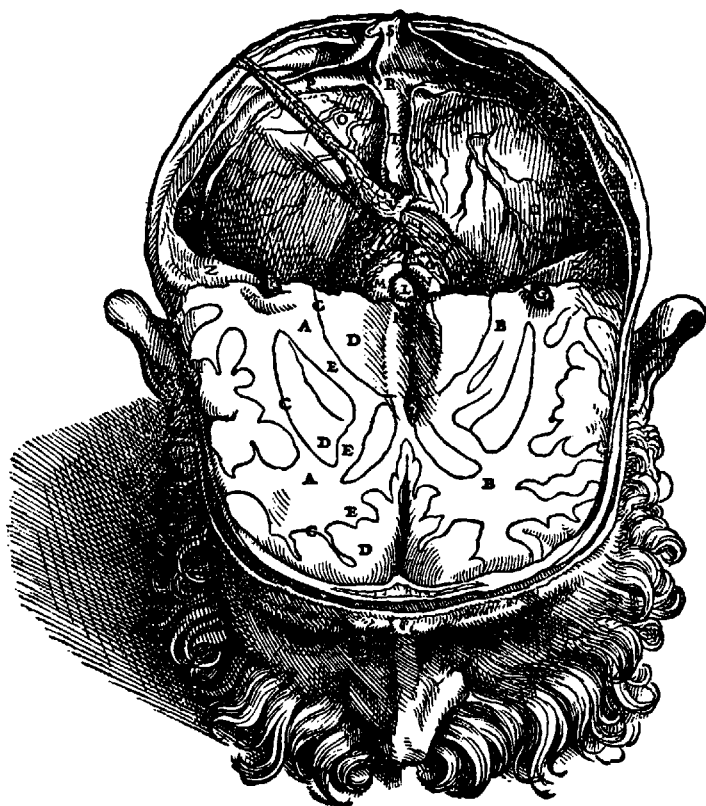


Figure 4.2

système nerveux juste au milieu de la partie arrière du cerveau. Comme il est tout à fait impossible de discerner quelle est sa fonction (on ne voit d'ailleurs toujours pas à quoi sert la glande pinéale), Descartes a proposé d'en faire le fondement même de la conscience : d'après lui, elle serait la gare où arrive la circulation venue des sens, la cause de la transaction magique entre le cerveau matériel d'une personne et son esprit immatériel... Selon Descartes, toutes les réactions corporelles ne requièrent pas cette intervention de l'esprit conscient. Il connaissait bien l'existence de ce que l'on appelle aujourd'hui des réflexes, et il postulait qu'ils devaient résulter de courts-circuits entièrement mécaniques traversant la gare pinéale et par conséquent inconscients.

Descartes avait tort sur les détails. Il pensait en effet que le feu déplaçait la peau, qui tirait un fil minuscule, qui ouvrait un pore dans le ventricule (F) qui provoquait le flux d'un « esprit animal » à travers un tube creux, lequel faisait enfler les muscles



Figure 4.3

de la jambe et causait le retrait du pied (Descartes, 1664). Mais à part ça, l'idée était bonne. On ne peut pas en dire autant de la conception cartésienne de la glande pinéale comme plaque tournante de la conscience de ce que nous pouvons appeler le goulot cartésien. Cette idée, qui est la base du dualisme cartésien, est désespérément fausse, comme nous l'avons vu au chapitre 1. Mais alors que le matérialisme quelle que soit sa forme est aujourd'hui une opinion reçue presque par tout le monde, même les matérialistes les plus subtils d'aujourd'hui oublient souvent qu'à partir du moment où l'on rejette la fantomatique *res cogitans* de Descartes, il n'est plus nécessaire de postuler une porte d'entrée centrale, ni même un quelconque centre *fonctionnel* du cerveau. Non seulement la glande pinéale n'est pas la machine à télécopie du cerveau, mais elle n'est pas non plus le palais de l'Élysée² du cerveau, pas plus que d'autres portions du cerveau. Le cerveau est le Quartier Général, l'endroit où se trouve l'ultime observateur, mais il n'existe pas de raison de croire que le cerveau lui-même ait un quartier général plus profond, un sanctuaire intérieur, et que le fait qu'on puisse arriver à ce sanctuaire soit une condition nécessaire et suffisante de l'expérience consciente. En d'autres termes, il n'y a pas d'observateur à l'intérieur du cerveau³.

La lumière voyage bien plus vite que le son, comme l'exemple des feux d'artifice nous le rappelle. Mais nous savons à présent que le cerveau met plus de temps à traiter les stimuli visuels que les stimuli auditifs. Comme le neuropsychologue Ernst Pöppel (1985, 1988) l'a fait remarquer, en raison de ces différences de contrepoids, l'« horizon de simultanéité » est d'*environ* dix mètres : la lumière et le son qui quittent le même point situé à environ dix mètres des organes sensoriels de l'observateur produisent des réponses neuronales qui sont « disponibles au centre » au même moment. Pouvons-nous être plus précis ? Cela pose un problème. Il ne s'agit pas seulement de mesurer les distances entre l'événement extérieur et les organes sensoriels ou les vitesses de transmission à l'intérieur des divers milieux, ou d'autoriser des différences particulières. Le problème le plus fondamental est celui de la détermination de la « ligne d'arrivée » dans le cerveau. Pöppel a obtenu ce résultat en comparant les mesures comportementales, les temps moyens de réaction (pousser sur le bouton) à des stimuli auditifs et visuels. La différence varie entre 30 msec et 40 msec, ce qui correspond au temps que met le son pour

faire à peu près 10 mètres (le temps que met la lumière à parcourir 10 mètres ne diffère pratiquement pas de zéro). Pöppel a utilisé une ligne d'arrivée périphérique – le comportement externe – mais nous avons naturellement l'intuition que *l'expérience* de la lumière ou du son se produit *entre* le moment où les vibrations touchent les organes de nos sens et celui où nous trouvons le moyen de pousser le bouton pour signaler cette expérience. Et cela arrive quelque part *au centre*, quelque part dans le cerveau sur les chemins neuronaux excités entre l'organe des sens et le doigt. Il semble que si nous pouvions dire *exactement* où, nous pourrions dire où a eu lieu l'expérience. Et *vice versa* : si nous pouvions dire exactement quand elle s'est produite, nous pourrions dire où dans le cerveau l'expérience consciente était localisée.

Appelons *matérialisme cartésien* l'idée selon laquelle il existe un lieu central de ce genre dans le cerveau : car c'est la thèse à laquelle vous êtes conduit quand vous abandonnez le dualisme de Descartes mais pas l'image d'un Théâtre central (mais matériel) où « tout se réunit ». La glande pinéale serait un bon candidat pour jouer le rôle d'un tel Théâtre Cartésien, mais on en a suggéré d'autres, le cingulum antérieur, la formation réticulée, divers lieux dans les lobes frontaux. Le matérialisme cartésien est la thèse selon laquelle il existe une ligne d'arrivée cruciale ou une frontière quelque part dans le cerveau, qui marque un endroit où l'ordre d'arrivée est identique à l'ordre de « présentation » dans le cerveau dans l'expérience parce que *ce qui se passe là* est ce dont vous êtes conscient. Il se peut que personne n'endosse explicitement le matérialisme cartésien. Beaucoup de théoriciens insisteraient sur le fait qu'ils ont rejeté explicitement une idée aussi mauvaise. Mais, comme nous le verrons, l'image séduisante du Théâtre Cartésien revient nous hanter – que nous soyons profanes non scientifiques – même une fois que nous avons dénoncé et exorcisé ce dualisme fantomatique.

Le Théâtre Cartésien est une représentation métaphorique de la façon dont l'expérience consciente doit avoir son siège dans le cerveau. Il semble de prime abord être une extrapolation à partir d'un fait familier et indéniable : *pour des intervalles de temps macroscopique usuels*, nous pouvons bien sérier les événements en deux catégories, celle des événements qui ne sont « pas encore observés » et celle des événements « qui restent à observer ». Nous le faisons en localisant l'observateur en un certain point pour

ensuite disposer les mouvements des véhicules d'information relativement à ce point. Mais quand nous essayons d'étendre cette méthode de manière à expliquer les phénomènes qui impliquent des intervalles de temps très courts, nous rencontrons une difficulté *logique* : si le « point » de vue de l'observateur doit s'étaler sur un volume assez vaste du cerveau de l'observateur, la saisie subjective que doit avoir celui-ci de la suite et de la simultanéité *doit* être déterminée par quelque chose d'autre que « l'ordre d'arrivée », parce que l'ordre d'arrivée n'est pas défini complètement tant que la destination pertinente n'est pas spécifiée. Si A bat B sur la ligne d'arrivée, mais si B bat A sur une autre ligne d'arrivée, quel est le résultat qui fixe la séquence subjective dans la conscience ? (Minsky, 1985, p. 61). Pöppel parle des moments où la vue et le son deviennent « disponibles centralement » dans le cerveau. Mais quel est le point ou quels sont les points de « disponibilité centrale » qui doivent compter comme déterminants d'un ordre *dont on fait l'expérience*, et pourquoi ? Quand nous essaierons de répondre à cette question, nous serons forcés d'abandonner le Théâtre Cartésien et de le remplacer par un nouveau modèle.

L'idée qu'il existerait un centre spécial dans le cerveau est la plus mauvaise et la plus tenace de toutes les idées qui empoisonnent nos modes de pensée au sujet de la conscience. Comme nous le verrons, elle revient sans cesse sous de nouveaux déguisements, et on l'invoque pour diverses raisons qui semblent irrésistibles. Tout d'abord, on invoque notre appréciation personnelle et introspective de « l'unité de la conscience », qui nous suggère de distinguer entre « le dedans » et « le dehors ». La frontière que nous établissons naïvement entre le « moi » et le « monde extérieur » est celle de notre peau (et des lentilles de nos yeux), mais au fur et à mesure que nous en apprenons plus sur la façon dont les événements qui se produisent dans notre corps peuvent « nous » être inaccessibles, le grand dehors s'introduit sans cesse. « Là dedans », je peux essayer de lever mon bras, mais « là dehors », il « s'est endormi » ou il est paralysé et il ne bougera pas ; mes lignes de communication à partir de l'endroit où *je* peux me trouver dans la machinerie neuronale qui contrôle mon bras ont été corrompues. Et si mon nerf optique était coupé d'une façon ou d'une autre, je ne m'attendrais pas à continuer à voir même si mes yeux étaient encore intacts ; avoir des expériences visuelles

se produit apparemment à l'intérieur de mes yeux, quelque part entre mes yeux et ma voix quand je vous dis ce que je vois.

Ne s'ensuit-il pas qu'en vertu d'une nécessité géométrique, nos esprits conscients sont localisés au *point terminal* de tous les processus *internes*, juste avant que s'initient tous les processus *externes* qui réalisent nos actions ? En avançant de la périphérie le long des canaux d'entrée d'information qui partent de l'œil, par exemple, nous remontons le nerf optique, puis diverses zones du cortex visuel. Et puis, où allons-nous donc ? En avançant à partir de l'autre périphérie, en nageant à contre-courant, à partir des muscles et des neurones moteurs qui les contrôlent, nous arrivons à l'aire motrice supplémentaire dans le cortex. Et puis, où allons-nous donc ? Ces deux voyages convergent suivant deux tracés, le tracé afférent (entrée) et le tracé efférent (sortie). Aussi difficile que soit la tâche consistant à déterminer en pratique la localisation précise de la « Division Continentale » dans le cerveau, ne doit-il pas exister, du fait d'une pure extrapolation géométrique, un point culminant, un point focal, tel que toutes les lacunes qui se situent d'un côté de ce point sont *pré-expérientielles*, alors que toutes les lacunes qui sont de l'autre côté sont *post-expérientielles* ?

Selon l'image cartésienne, c'est une évidence qui s'impose à l'introspection visuelle, parce que tout s'engouffre vers la gare constituée par la glande pinéale et à partir d'elle. Il semblerait donc que si nous devions adopter un modèle plus habituel des processus du cerveau, nous devrions être capables de coder nos explorations avec des couleurs, en utilisant, par exemple, le rouge pour ce qui est afférent et le vert pour ce qui est efférent ; chaque fois que nos couleurs changeraient soudainement, nous aurions un point médiant fonctionnel dans la grande Division Mentale.

Cet argument si curieusement séduisant en évoque un autre. Il recoupe un argument fallacieux qui a eu beaucoup trop d'influence ces derniers temps : la célèbre courbe d'Arthur Laffer, fondement intellectuel de la Reaganéconomie (pour aller vite). Si le gouvernement impose à 0 %, il n'a aucun revenu ; s'il impose à 100 %, personne ne travaillera pour gagner de l'argent ; par conséquent, le gouvernement n'aura pas de revenus ; à 2 %, il aura en gros deux fois le revenu qu'il obtient à 1 %, etc. Au fur et à mesure que le taux d'imposition augmente, on verra arriver des rendements décroissants : les impôts deviendront onéreux. Si l'on regarde de l'autre côté de l'échelle, 99 % de taux d'im-

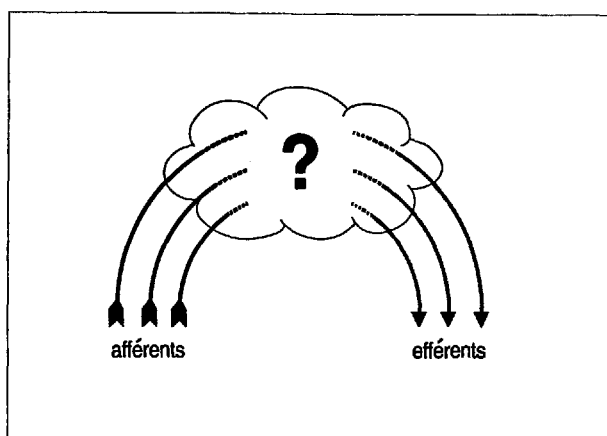


Figure 4.4

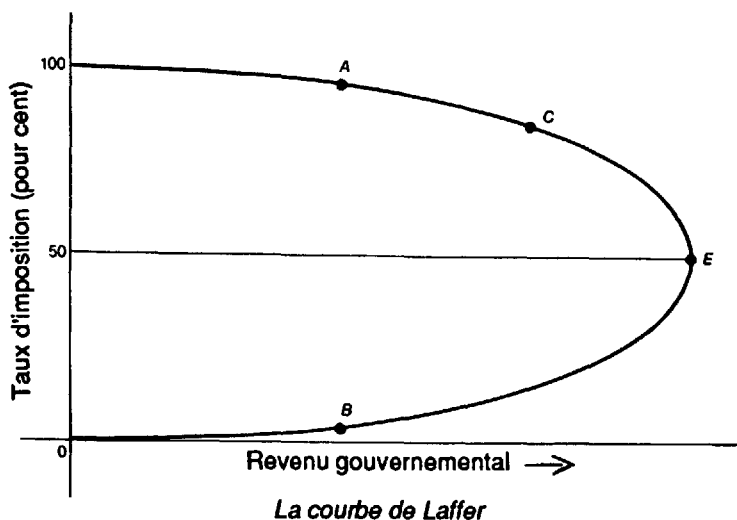
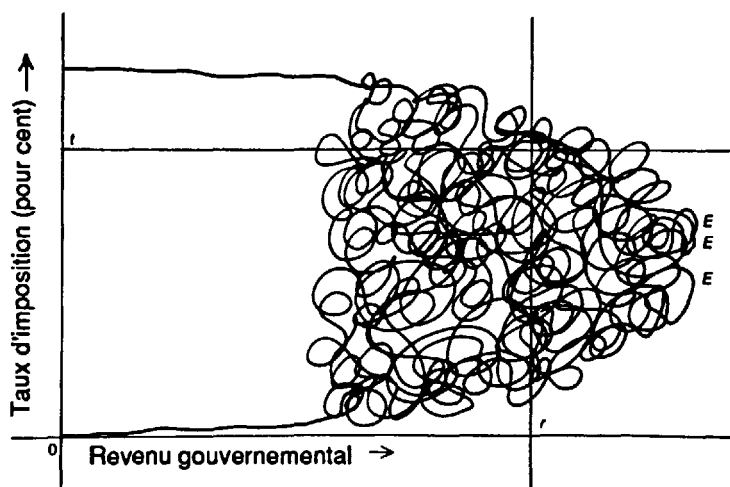


Figure 4.5

De « Mathematical Games », par Martin Gardner, dessin de Ilil Arbel, © *Scientific American*, décembre 1981.

position est à peine moins satisfaisant que 100 %, en sorte qu'il n'y aura presque pas d'augmentation du revenu ; à 90 % le gouvernement fera mieux, et encore mieux au taux plus tentant de 80 %. Les tracés particuliers de la courbe qui apparaissent sur la figure peuvent ne pas être symétriques. Mais ne doit-il pas exister, en vertu d'une nécessité géométrique, un endroit où la courbe s'infléchit, un taux de taxation qui maximise le revenu ?

L'idée de Laffer a consisté à dire que puisque le taux actuel d'imposition était situé sur la partie supérieure de la courbe, abaisser les impôts augmenterait en fait les revenus. C'était une idée tentante ; et beaucoup de gens ont jugé qu'elle devait être correcte. Mais comme Martin Gardner l'a remarqué, du fait même que les points extrêmes de la courbe sont clairs, il n'y a pas de raison de penser que la partie inconnue de la courbe dans les régions médianes doive avoir l'air continue. Dans un esprit satirique, il propose une courbe rivale, la « courbe néo-Laffer », qui a plus d'un « maximum », et l'accessibilité de l'un ou l'autre de ces maxima dépend des complexités de l'histoire et des circonstances, et non d'un changement d'une seule variable (Gardner, 1981). Nous pourrions tirer la même morale quant à ce qui se passe dans les brumes situées derrière les zones périphériques afférentes et efférentes : la clarté des zones périphériques ne nous garantit en rien que la même distinction continuera à s'appliquer *partout à l'intérieur*. Le « grognement technique » que Gardner envisage pour l'économie est clair comme de l'eau de roche comparé au grouillement des activités qui se produisent dans les régions les plus centrales du cerveau. Nous devons cesser de penser que le cerveau peut avoir un sommet fonctionnel ou point central. Ce n'est pas là un raccourci ano-



La courbe néo-Laffer (GN)

Figure 4.6

De « Mathematical Games », par Martin Gardner, dessin de Ilil Arbel, © *Scientific American*, décembre 1981.

din : c'est une mauvaise habitude. Pour rompre avec cette mauvaise habitude de pensée, il nous faut examiner quelques cas dans lesquels elle se manifeste, mais il nous faut aussi la remplacer par une image rivale.

Introduction du modèle des Versions Multiples

Voici un premier substitut, le modèle des Versions Multiples pour la conscience. Il pourra paraître étrange et difficile d'accès à première vue. Cela montre à quel point l'idée du Théâtre Cartésien est tenace. Selon le modèle des Versions Multiples, toutes les espèces de perceptions – en fait toutes les espèces de pensées et d'activité mentales – sont traitées dans le cerveau par des processus parallèles et multiples d'interprétation et d'élaboration des entrées sensorielles. L'information qui pénètre dans le système nerveux fait l'objet d'une « révision éditoriale » continue. Par exemple, pendant que votre tête bouge légèrement et que vos yeux bougent beaucoup, les images sur vos rétines se déplacent constamment, un peu comme celles des films d'amateurs prises par des gens qui ne tiennent pas bien la caméra. Mais ce n'est pas la façon dont les choses nous apparaissent. Les gens sont souvent surpris d'apprendre que dans les conditions normales, leurs yeux dansent en saccades rapides, avec environ cinq fixations rapides par seconde, et que ce mouvement, comme le mouvement de leur tête, est édité très tôt dans le processus qui va du globe oculaire à... la conscience. Les psychologues en ont appris beaucoup sur les mécanismes qui permettent de parvenir à ces effets normaux, et ils ont aussi découvert certains effets spécifiques, comme l'interprétation de la profondeur par des stéréogrammes de points distribués au hasard (Julesz, 1971) (voir figure 4.7, p. 147).

Si vous regardez ces deux carrés légèrement différents à travers un stéréoscope (ou si vous les contemplez en louchant légèrement de manière à ce que les deux images se fondent en une seule – certains peuvent y parvenir sans dispositif stéréoscopique), vous finirez par voir émerger une forme tri-dimensionnelle. Le cerveau met en effet en œuvre un processus éditorial impressionnant qui compare l'information venue de chaque œil. On peut parvenir à trouver l'enregistrement globalement optimal sans avoir d'abord à soumettre toutes les données à un processus élaboré d'extraction



Figure 4.7

d'information. À un niveau inférieur, les saillances – les points individuels dans un stéréogramme de points aux hasards – comportent suffisamment de coïncidences pour dicter une solution ⁴.

Les processus d'édition du cerveau mettent un bon moment pour produire ces effets, mais il en existe d'autres qui sont plus rapides. L'effet Mc Gurk (Mc Gurk et Mc Donald, 1979) en est un exemple. Quand un film en français est sous-titré en anglais, la plupart du temps, les spectateurs ne réalisent pas le décalage entre les mouvements des lèvres qu'ils voient et les sons qu'ils entendent – sauf si le sous-titrage est mal fait. Que se passe-t-il si on fabrique une bande-son qui s'accorde bien avec les images sauf pour quelques consonnes délibérément non accordées ? (En utilisant notre vieille connaissance à nouveau, nous pouvons supposer que les lèvres de la personne filmée disent « de gauche à droite » et que la bande-son dise « de bauche à loite ».) Quelle va être l'expérience éprouvée par les gens ? Ils entendront « de gauche à droite ». Dans la lutte éditoriale entre les contributions qui viennent des yeux et des oreilles, les yeux gagnent – dans ce cas au moins.

Ces processus éditoriaux se produisent durant d'importantes fractions de seconde, pendant lesquelles diverses additions, incorporations, amendements et réécritures de contenu peuvent survenir, dans des ordres variés. Nous ne faisons pas directement l'expérience de ce qui se passe dans nos rétines, dans nos oreilles, sur la surface de notre peau. Ce dont nous faisons effectivement l'expérience est le produit de nombreux processus d'interprétation

– des processus éditoriaux, en fait. Ils traitent des représentations relativement brutes et grossières, et ils prennent place dans des flux d'activité qui se déroulent dans diverses parties du cerveau. La plupart des théories de la perception le reconnaissent. À présent, nous pouvons ajouter une nouvelle caractéristique du modèle des Versions Multiples : les détections de traits *n'ont besoin d'être faites qu'une seule fois*. Cela veut dire qu'à partir du moment où une « observation » particulière d'un trait quelconque a été faite par une portion spécialisée et localisée du cerveau, le contenu d'information ainsi fixé n'a pas besoin d'être envoyé à un autre endroit pour être rediscriminé par un quelconque « discriminateur » en chef. En d'autres termes, la discrimination ne conduit pas à une *représentation* du trait déjà discriminé au bénéfice de l'auditoire du Théâtre Cartésien – car il n'y a pas de Théâtre Cartésien.

Ces fixations de contenu spatialement et temporellement distribuées peuvent être précisément localisées à la fois dans l'espace et dans le temps, mais leurs surgissements ne sont pas la marque du surgissement de la conscience de leurs contenus. La question de savoir si un contenu particulier ainsi discriminé apparaîtra comme un élément de l'expérience consciente est toujours ouverte, et, comme nous le verrons, de demander *quand il devient conscient* est une confusion. Ces discriminations de contenus distribuées produisent, au bout d'un moment, quelque chose *qui ressemble assez* à un flux narratif ou à une séquence. On peut considérer cette dernière comme sujette à un processus d'édition continu effectué par de nombreux processus qui sont distribués à la ronde dans le cerveau et se poursuivent indéfiniment dans le futur. Ce flux de contenus est semblable à celui d'une narration seulement parce qu'il est multiple : à tout moment, il existe de multiples « versions » de fragments narratifs qui surviennent à divers moments de l'édition et à différents endroits du cerveau.

Sondé à différents endroits et à différents moments, ce flux produit divers effets et provoque différents récits de la part du sujet. Si l'on diffère trop ce sondage (pendant la nuit, par exemple), le résultat risque de ne plus produire une narration – ou de produire une narration qui a été digérée ou « rationnellement reconstruite » jusqu'à perdre son intégrité. Si l'on sonde « trop tôt », on peut regrouper des données sur l'origine d'une discrimination particulière dans le cerveau, mais on risque alors de dévier ce qui aurait été autrement la progression normale du flux multiple.

Mieux encore, le modèle des Versions Multiples évite l'erreur si tentante qui consiste à supposer qu'il doit y avoir un récit unique (la version « finale » ou « publiée », pourrait-on dire) qui soit canonique – et qui soit le flux *réel* de la conscience du sujet, que l'expérimentateur (ou même le sujet) ait ou non accès à celle-ci.

En l'état actuel des choses, ce modèle n'a sans doute pas grand sens pour vous en tant que modèle explicatif de la conscience telle que vous la connaissez à partir de votre expérience privée. Cela tient au fait que vous continuez trop aisément à croire que votre conscience se déroule dans le Théâtre Cartésien. Rompre avec cette habitude naturelle et confortable pour en venir à considérer le modèle des Versions Multiples comme un rival plausible prendra du temps et demandera des efforts qui seront loin d'être faciles. Cela sera sûrement la partie la plus difficile du livre, mais elle est essentielle pour l'ensemble de la théorie et on ne peut s'en dispenser ! On ne trouvera pas de mathématiques dans ces pages, Dieu merci. Il suffira de réfléchir avec un peu d'attention et de clarté, en s'assurant que l'on se représente bien les choses, sans se laisser séduire par de fausses images. Tout au long de ce chemin semé d'embûches, nous ferons diverses expériences de pensée pour stimuler notre imagination. Préparons-nous donc à des exercices assez ardu. Au bout du compte, c'est à une nouvelle conception de la conscience que l'on parviendra, entraînant une réforme profonde (mais pas une révolution radicale) de nos modes de pensée concernant le cerveau. (Un modèle similaire a été proposé par William Calvin sous le nom de « tissage de scénarios ».)

Il existe un bon moyen pour comprendre une nouvelle théorie : il suffit de se demander comment elle rend compte d'un phénomène relativement simple qui n'est pas expliqué par l'ancienne théorie. Le dispositif A est une découverte sur le mouvement apparent qui a été provoquée, je suis heureux de le signaler, par la question d'un philosophe. Pour obtenir des images de cinéma et de télévision, on crée des mouvements apparents en présentant une succession rapide d'images « statiques ». Depuis les tout débuts du cinéma, les psychologues ont étudié ce phénomène appelé *phi* par Wertheimer (1912), qui fut le premier à l'étudier systématiquement. Dans le cas le plus simple, si l'on allume successivement et rapidement deux ou plusieurs petits points séparés par au plus quatre degrés d'angle visuel, on aura l'impression qu'un seul point va et vient. Plusieurs variantes de *phi* ont été étudiées. L'une des plus frappantes est rapportée par les psycho-

logues Paul Kolers et Michael von Grünau (1976). Le philosophe Nelson Goodman a demandé à Kolers si le phénomène phi persistait si les deux points lumineux étaient de couleur différente et ce qui arrivait ensuite à la couleur « du » point pendant qu'« il » se déplaçait. L'illusion de mouvement disparaîtrait-elle et serait-elle remplacée par deux points s'illuminant séparément ? Est-ce qu'un point « se mouvant » illusoirement changerait graduellement de couleur, traçant une trajectoire à travers le solide des couleurs (la sphère tridimensionnelle qui distribue toutes les teintes) ? (Rien ne vous empêche de faire votre propre prédiction avant de lire la suite.) La réponse, quand Kolers et von Grünau ont fait les expériences, était inattendue : deux points colorés différents s'allumèrent pendant 150 msec chacun (avec un intervalle de 50 msec) ; le premier point parut commencer à se mouvoir et à changer brusquement de couleur *au milieu de son passage illusoire* vers la seconde localisation. Goodman s'est demandé « Comment nous sommes capables [...] d'intercaler le point aux moments et lieux intermédiaires sur un chemin qui va du premier au second *avant que cette seconde lumière ne s'allume* » (Goodman, 1978, p. 73).

On peut évidemment poser la même question pour n'importe quel phénomène phi, mais le phénomène phi de couleur étudié par Kolers est particulièrement révélateur. Supposez que le premier point soit rouge et que le second, déplacé, soit vert. À moins qu'il n'y ait « précognition » dans le cerveau (hypothèse extravagante que nous différerons indéfiniment), le contenu illusoire, *rouge-passant-au-vert-en-milieu-de-course*, ne peut pas être créé avant un moment *ultérieur* où un second point, vert, est identifié comme advenant dans le cerveau. Mais si le second point est déjà « dans l'expérience consciente », ne sera-t-il pas trop tard pour interposer le contenu illusoire entre l'expérience consciente de la tache rouge et l'expérience consciente de la tache verte ? Comment le cerveau réalise-t-il ce tour de passe-passe ?

Le principe selon lequel les causes doivent précéder leurs effets s'applique aux multiples processus distribués qui accomplissent le travail d'édition dans le cerveau. Tout processus particulier requérant de l'information à partir d'une source quelconque doit bien entendu attendre de recevoir cette information : il ne peut aller ici tant qu'il n'est pas allé là. C'est ce qui interdit les explications « magiques » ou précognitives du phénomène phi avec les changements de couleur. Le contenu *point vert* ne peut

être attribué à un événement quelconque, conscient ou inconscient, tant que la lumière venue du point vert n'a pas atteint l'œil et déclenché l'activité neurale normale dans le système jusqu'au niveau où la discrimination du vert se réalise. Ainsi, la discrimination (illusoire) du vert-devenant-rouge doit s'accomplir *après* la discrimination du point vert. Mais puisque votre expérience consciente est *d'abord vert, ensuite vert-devenant-rouge, puis enfin vert*, il s'ensuit (« sûrement ») que votre conscience de l'ensemble de l'événement doit être retardée jusqu'à un moment qui suit celui où le point vert est (inconsciemment ?) perçu. Si vous trouvez cette conclusion convaincante, vous êtes toujours aux prises avec le Théâtre Cartésien. Une expérience de pensée vous permettra de lui échapper.

Les révisions orwelliennes et staliniennes

Je ne sais pas vraiment si les autres me perçoivent ou si, lorsque, durant une fraction de seconde, mon visage interfère avec leur horizon, un millionième de seconde après qu'ils aient porté leur regard sur moi, ils commencent déjà à me tirer de leur mémoire : oublié avant d'arriver à l'archange faible et triste d'un souvenir.

Ariel Dorfman, *Mascara*, 1988.

Supposez que je trafique votre cerveau pour insérer dans votre mémoire une femme factice portant un chapeau par exemple à la soirée où vous étiez dimanche dernier alors qu'il n'y en avait pas. Si lundi, quand vous vous rappelez la soirée, vous vous souvenez d'elle et ne pouvez pas trouver de ressources internes vous permettant de douter de la véracité de votre mémoire, nous n'en dirons pas moins que vous ne l'avez pas *vraiment* vue à la soirée de dimanche. Bien sûr, les souvenirs (factices) dont vous avez ensuite fait l'expérience peuvent être aussi nets que possible, et le mardi venu nous pouvons certainement admettre que vous avez l'expérience consciente précise d'avoir vu une femme portant un chapeau à la soirée, mais la *première* expérience de ce genre que vous ayez eu, dites-vous avec insistance, se déroulait lundi, pas dimanche (bien qu'il ne vous semble pas qu'il en soit ainsi).

Nous n'avons pas la possibilité d'insérer des souvenirs factices

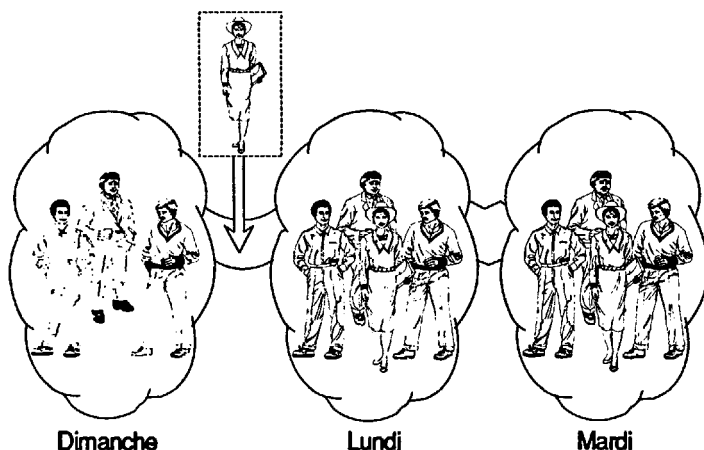


Figure 4.8

par neurochirurgie, mais il arrive que nos souvenirs nous jouent des tours. Ce que nous ne pouvons pas réussir par chirurgie arrive tout seul dans le cerveau. Quelquefois, nous avons l'impression de nous souvenir, même nettement, d'expériences qui n'ont jamais eu lieu. Appelons *orwelliennes*, de telles contaminations ou révisions de mémoire d'après la vision à vous donner froid dans le dos que George Orwell donne dans le roman *1984*, où le ministère de la Vérité, s'occupe de réécrire l'histoire et refuse l'accès au passé (réel) aux générations suivantes.

Le fait que des révisions (orwelliennes) postérieures à l'expérience soient possibles révèle un aspect de l'une de nos distinctions fondamentales : la distinction entre l'apparence et la réalité. Parce que nous reconnaissons la possibilité (au moins en principe) de révisions orwelliennes, nous reconnaissons que nous risquons d'inférer à « c'est ce qui s'est réellement passé » de « c'est ce dont je me souviens ». Par conséquent, nous résistons – à juste titre – à l'« opérationnalisme » diabolique qui essaie de nous convaincre que ce dont nous nous souvenons (ou ce que l'histoire rapporte dans les archives) est *précisément* ce qui s'est vraiment passé ⁵.

La révision orwellienne est l'une des façons possible de tromper la postérité. Il en existe une autre : elle consiste à monter de faux procès, à mettre en forme des faux témoignages et des confessions factices, pleines de preuves simulées. Appelons cela la tactique *stalinienne*. Si nous sommes en général sûrs que c'est

de telle ou telle sorte de falsification que nous sommes victimes, orwellienne ou stalinienne, ce n'est que par un accident heureux. Dans toute campagne réussie de désinformation, si nous nous demandions si les faits rapportés par les journaux étaient des comptes rendus orwelliens de procès qui ne se sont jamais produits ou des comptes rendus corrects de procès bidons qui ont effectivement eu lieu, nous serions incapables de dire où est la différence. Si *toutes* les traces – journaux, cassettes vidéo, mémoires, inscriptions sur des pierres tombales, témoins vivants – ont été oblitérées ou révisées, nous n'aurions jamais aucun moyen de savoir s'il y eut *d'abord* fabrication et ensuite un procès public dont nous avons le récit devant les yeux, ou si, *après* une exécution sommaire, l'histoire a été fabriquée pour recouvrir ce qui s'est passé, de sorte qu'il n'y a jamais *réellement* eu de procès.

Dans la vie de tous les jours, la distinction entre les méthodes orwelliennes et les méthodes staliniennes pour produire des pseudo-archives fonctionne sans problème à des échelles de temps macroscopiques. Pourquoi ne s'appliquerait-elle pas sans problème *partout* ? En fait, c'est une illusion. Une expérience de pensée permet de la prendre sur le fait : cette expérience ne diffère de celle qui vient d'être considérée que du point de vue de l'échelle temporelle.

Supposez que vous vous teniez à un coin de rue et qu'une femme aux longs cheveux vienne à passer. À peu près une seconde *après* cet événement, un souvenir sous-jacent d'une femme antérieure – une femme aux cheveux courts avec des lunettes – contamine le souvenir de ce que vous venez de voir : quand on vous demande une minute plus tard des détails sur la femme que vous venez de voir, vous rapportez, sincèrement mais en vous trompant, qu'elle avait des lunettes. Tout comme dans le cas de la femme au chapeau durant la soirée, vous êtes enclin à dire que votre expérience *visuelle* originelle, par opposition au souvenir intervenu quelques secondes plus tard, n'était *pas* l'expérience d'une femme portant des lunettes. Mais en raison de la contamination ultérieure de vos souvenirs, tout se passe comme si au premier moment où vous l'avez vue, vous aviez été frappé par ses lunettes. Une révision orwellienne est intervenue : durant un bref instant, avant la contamination du souvenir, il ne vous a *pas* semblé qu'elle portait des lunettes. Pendant ce bref moment, la *réalité* de votre expérience consciente était celle d'une femme aux cheveux longs *sans* lunettes, mais

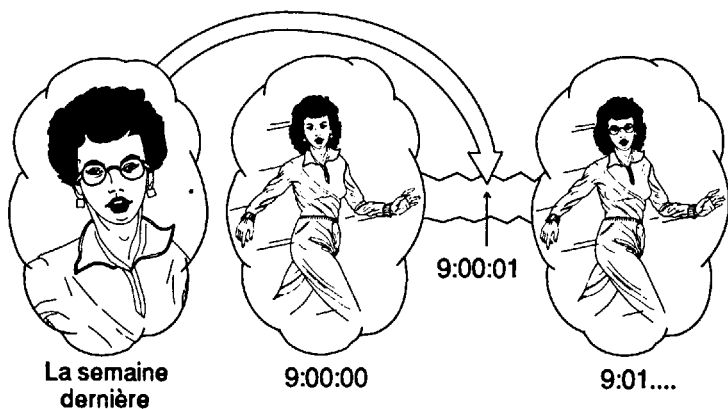


Figure 4.9

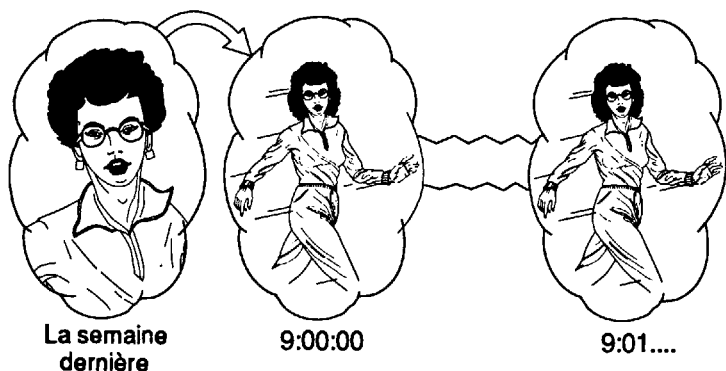


Figure 4.10

ce fait historique est devenu inerte ; il n'a pas laissé de trace, en raison de la contamination de la mémoire qui s'est produite une seconde après que vous l'ayez vue.

Cette analyse de ce qui s'est produit se trouve cependant menacée par une analyse rivale. Les souvenirs antérieurs sous-jacents que vous avez de la femme aux lunettes auraient aussi bien pu contaminer *en amont* votre expérience, dans le traitement de l'information intervenant « avant la conscience », en sorte que vous fassiez *une expérience hallucinatoire* des lunettes dès le début de votre expérience. Dans ce cas, votre souvenir obsessionnel de la femme antérieure avec ses lunettes vous jouerait un tour

stalinien : il créerait une sorte de faux procès dans l'expérience, dont vous vous souviendriez ensuite nettement à des moments ultérieurs, grâce à l'enregistrement dans votre mémoire. Pour le sens commun, ces deux cas sont aussi différents que possible. Dans le premier cas (figure 4.9), vous n'êtes pas victime d'une hallucination au moment où la femme passe, mais vous êtes victime ultérieurement d'une hallucination mémorielle : vous avez de faux souvenirs de votre expérience « réelle ». Dans le second (figure 4.10), vous avez une hallucination quand elle passe en courant, et à partir de là vous avez ensuite un souvenir net de cette hallucination (qui « a eu réellement lieu au sein de votre conscience »). Ces possibilités ne sont-elles pas différentes, quel que soit le découpage temporel que nous adoptons ?

Non. Ici, il n'est pas garanti que la distinction entre les révisions perceptives et les révisions mémorielles qui fonctionnent si bien à d'autres échelles ait un sens. Nous avons pénétré dans la zone brumeuse dans laquelle le point de vue du sujet est spatialement et temporellement estompé. La question de savoir s'il s'agit d'une falsification orwellienne ou stalinienne perd de son intérêt.

Il existe une fenêtre temporelle qui a commencé avec la femme aux cheveux longs qui passait, excitant vos rétines, et qui s'est terminée quand vous avez exprimé – à vous-même ou à quelqu'un d'autre – votre conviction finale qu'elle portait des lunettes. À un certain moment pendant cet intervalle, le contenu *portant des lunettes* a été ajouté de manière erronée au contenu *femme aux cheveux longs*. Nous pouvons supposer (et pourrions finalement le confirmer en détail) qu'il y a eu un bref moment pendant lequel le contenu *femme aux cheveux longs* avait déjà été discriminé dans le cerveau mais *avant* que le contenu *portant des lunettes* ait été faussement « lié » à lui. Bien entendu, il serait raisonnable de supposer que c'est cette discrimination d'une femme à cheveux longs qui a déclenché le souvenir de la femme antérieure portant les lunettes. Mais ce que nous ne saurions pas, c'est si cette liaison erronée a eu lieu « avant ou après le fait », le fait présumé « d'expérience consciente effective ». Avez-vous d'abord été conscient d'une femme aux cheveux longs sans lunettes et ensuite d'une femme aux cheveux longs avec des lunettes, état conscient ultérieur qui a balayé le souvenir de l'expérience antérieure, ou bien le tout premier moment d'expérience consciente a-t-il été faussement teinté de lunettes ?

Si le matérialisme cartésien était vrai, cette question devrait avoir une réponse, même si on ne pouvait la soumettre rétrospectivement à un test empirique. Car le contenu « qui a franchi le premier la ligne d'arrivée » était soit *femme aux cheveux longs* soit *femme aux cheveux longs avec des lunettes*. Mais presque tous les théoriciens insisteraient pour dire que le matérialisme cartésien est faux. Cependant, ils n'ont pas vu que cela impliquait qu'il n'existe pas de ligne d'arrivée et que, par conséquent, l'ordre temporel des discriminations ne peut être ce qui fixe l'ordre subjectif dans l'expérience. Ce n'est pas une conclusion facile à admettre, mais nous pouvons la rendre plus convaincante en examinant les problèmes que vous allez rencontrer si vous en restez à la solution traditionnelle.

Considérez le phénomène phi pour les couleurs de Kolers. Les sujets *rapportent* qu'ils voient la couleur du point en mouvement passer du rouge au vert au milieu de sa trajectoire. Kolers a eu l'idée ingénieuse de préciser cette séquence de texte en utilisant un pointeur, que les sujets ont « superposé » « rétrospectivement-mais-aussi-vite-que-possible » sur la trajectoire du point illusoirement en mouvement : en plaçant le pointeur, ils ont effectué un acte de langage qui avait pour contenu l'énoncé « Le point a changé de couleur exactement *ici* » (Kolers et von Grünau, 1976, p. 330).

Par conséquent, dans le monde hétérophénoménologique des sujets, il existe un changement de couleur en milieu de trajectoire, et l'information au sujet de la couleur vers laquelle on passe (et la direction du mouvement) doit venir de quelque part. Rappelez-vous la manière dont Goodman exprime l'énigme : « Comment sommes-nous capables [...] d'intercaler le point aux moments et lieux intermédiaires sur un chemin qui va du premier au second *avant que cette seconde lumière ne s'allume ?* » Peut-être, comme l'ont pensé certains auteurs, l'information vient-elle d'une *expérience antérieure*. Peut-être, comme le chien de Pavlov qui s'attendait à avoir de la nourriture quand la clochette sonnait, ces sujets s'attendaient-ils à voir le second point chaque fois qu'ils voyaient le premier et par la force de l'habitude, se représentaient-ils en fait le passage en anticipant une information sur le cas en question. Mais cette hypothèse a été réfutée. Même au premier essai (c'est-à-dire sans qu'il y ait possibilité de conditionnement), les gens ont l'expérience du phénomène phi. De plus, dans les essais ultérieurs, la direction et la couleur du second

point peuvent être changées au hasard sans que l'effet disparaisse. Par conséquent, d'une manière ou d'une autre, l'information qui vient du second point (quant à sa couleur et sa localisation) doit être utilisée par le cerveau pour créer la version « éditée » que rapportent les sujets.

Considérez, en premier lieu, l'hypothèse selon laquelle il existe un mécanisme stalinien : dans la salle de rédaction du cerveau, localisée avant la conscience, il y a un retard, une boucle de jeu comme dans la bande-son des émissions de radio « en direct » ; ce retard donne aux censeurs présents dans la salle de contrôle quelques secondes pour effacer des obscénités avant d'émettre le signal. *Dans la salle de rédaction*, l'image A, celle du point rouge, arrive en premier ; puis, quand l'image B, celle du point vert, arrive, des images intermédiaires (C et D) peuvent être créées et intercalées dans le film (dans l'ordre A, C, D, B) pour être ensuite projetées dans le théâtre de la conscience. Au moment où le « produit final » arrive dans la conscience, il a déjà son insertion illusoire.

Selon l'autre hypothèse, celle du mécanisme orwellien, peu de temps après la conscience du premier point *et* du second point (sans aucune illusion de mouvement apparent), une sorte d'historien révisionniste, installé dans le poste de réception de la mémoire-bibliothèque du cerveau, remarque que dans ce cas l'histoire non enluminée n'est pas assez plausible ; il interprète les événements bruts, des rouges-suivis-de-vert, en confectionnant un récit au sujet du passage intermédiaire ; il le complète avec un changement de couleur à mi-course et installe son histoire, en incorporant ses gloses, et en intercalant C et D

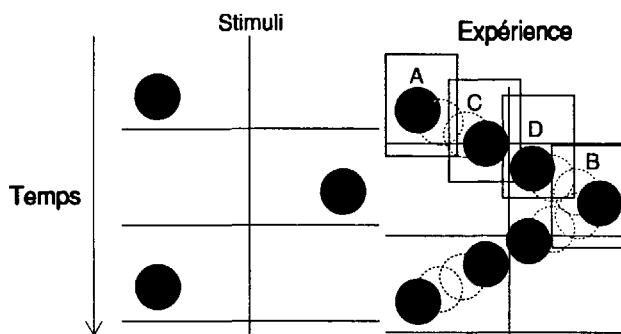


Figure 4.11

(figure 4.11) dans la mémoire-bibliothèque pour servir à toute référence future. Comme il travaille rapidement, en une fraction de seconde – le temps qu'il faut pour concevoir (mais pas pour énoncer) un compte rendu verbal de ce dont vous avez fait l'expérience – l'enregistrement auquel vous vous fiez, conservé dans la mémoire-bibliothèque, est déjà contaminé. Vous *dites* et vous *croyez* que vous avez vu le mouvement illusoire et le changement de couleur, mais c'est en fait une hallucination mémorielle, pas un souvenir exact de votre expérience consciente originelle.

Comment savoir laquelle de ces deux hypothèses est correcte ? Il pourrait sembler très simplement possible d'écarter l'hypothèse stalinienne, en raison du retard de conscience qu'elle postule. Dans l'expérience de Kolers et de von Grünau, une différence de 200 msec au départ sépare le point rouge et le point vert, et puisque, par hypothèse, *l'ensemble de l'expérience* ne peut pas être composée par la salle de rédaction avant que le contenu *point vert* ait atteint cette salle de rédaction, la conscience du point vert initial devra avoir été retardée d'au moins un temps similaire. (Si la salle de rédaction avait envoyé le contenu *point rouge* au théâtre de la conscience immédiatement, avant de recevoir l'image B et d'avoir fabriqué ensuite les images C et D, il est probable que le sujet ferait l'expérience d'une lacune dans le film, d'un retard d'au moins 200 msec entre A et C – aussi décelable qu'une lacune d'une syllabe dans un mot, ou de cinq images manquantes dans un film.)

Supposez que nous demandions aux sujets d'expérience de presser un bouton « dès qu'ils font l'expérience d'un point rouge ». Nous ne trouverions pas de différence – ou si peu – en temps de réponse face à un point rouge isolé par opposition à celle que produirait la présence d'un point rouge suivi 200 msec plus tard par un point vert (auquel cas les sujets rapportent un mouvement apparent de changement de couleur). Ne serait-ce pas parce qu'il y a *toujours* un retard d'au moins 200 msec dans la conscience ? Non. Il est bien établi que les réponses sous contrôle conscient, bien qu'elles soient plus lentes que des réponses comme des clignements d'yeux réflexes, surviennent avec des latences (retards) proches du minimum physiquement possible. Une fois soustrait les temps de parcours établi pour des pulsations entrant et sortant et le temps de préparation de la réponse, il ne reste pas assez de temps dans le « traitement central » pour cacher un

retard de 200 msec. Par conséquent, les réponses par pression sur le bouton doivent avoir été initiées avant la discrimination du second stimulus, le point vert.

On semble ainsi concéder la victoire à l'hypothèse orwellienne d'un mécanisme de révision postérieur à l'expérience : dès que le sujet devient conscient du point rouge, il se met à presser sur le bouton. *Pendant que s'enclenche la pression sur le bouton*, il devient conscient du point vert. *Alors*, ces deux expériences sont balayées de la mémoire et remplacées par l'enregistrement révisionniste du point rouge qui se meut et devient vert à mi-parcours. Le sujet rapporte immédiatement et sincèrement, *mais faussement*, qu'il a vu le point rouge se mouvoir vers le point vert avant de changer de couleur. S'il insiste pour dire qu'il était réellement conscient dès le départ du fait que le point rouge se mouvait et changeait de couleur, le théoricien orwellien lui expliquera fermement qu'il a tort ; sa mémoire lui joue des tours : le fait qu'il ait pressé le bouton au moment où il l'a fait montre bien qu'il était conscient du point rouge (stationnaire) avant que le point vert ne soit survenu. Après tout, ses instructions consistaient à presser le bouton *au moment où il avait conscience* d'un point rouge. Il doit avoir été conscient du point rouge environ 200 msec avant qu'il ait été conscient du fait qu'il se mouvait et devenait vert. Si ce n'est pas ainsi que les choses lui apparaissent, c'est simplement qu'il se trompe.

Le défenseur de l'hypothèse rivale stalinienne ne s'avouera cependant pas vaincu. En fait, insiste-t-il, le sujet a répondu à la tache rouge *avant* d'en être conscient ! Les directives qu'on lui a données (répondre à un point rouge) ont d'une certaine façon transpiré de la conscience vers la salle de rédaction, laquelle a provoqué (inconsciemment) l'action de pousser le bouton avant d'envoyer les versions éditées (les images ACDB) vers la conscience afin qu'elles soient « contemplées ». La mémoire du sujet ne lui a pas joué de tours : il rapporte exactement ce dont il était conscient, à cette nuance près qu'il insiste sur le fait qu'il a consciemment poussé le bouton après avoir vu le point rouge ; l'action « prématurée » de pousser le bouton a été déclenchée inconsciemment (ou préconsciemment).

La théorie stalinienne postule que le sujet a réagi en poussant le bouton à une détection inconsciente du point rouge ; la théorie orwellienne postule une expérience *consciente* d'un point rouge qui est immédiatement oblitérée dans la mémoire par ce qui

suit. Voilà le hic : nous sommes en présence de deux modèles différents de ce qui se passe dans le phénomène phi pour les couleurs. L'un postule un « bourrage de crâne » stalinien qui a lieu sur le chemin en amont, avant l'expérience, et l'autre une « révision mémorielle » orwellienne sur le chemin en aval. *Tous deux* sont compatibles avec *tout* ce que le sujet pense ou se rappelle. Notez que l'incapacité où nous sommes de distinguer ces deux possibilités ne s'applique pas seulement aux *observateurs extérieurs* dont on pourrait supposer qu'ils n'ont pas accès aux données privées auxquelles le sujet lui-même aurait un « accès privilégié ». Vous, en tant que sujet dans une expérience sur le phénomène phi, *ne pourriez pas* découvrir quoi que ce soit vous permettant de pencher en faveur de l'une des théories plutôt que l'autre ; vous éprouveriez « la même » expérience dans un cas comme dans l'autre.

Est-ce réellement le cas ? Que se passerait-il si vous étiez très attentif à votre expérience ? Ne seriez-vous pas capable de dire quelle est la différence ? Supposez que l'expérimentateur vous facilite la tâche, en ralentissant le dispositif et en allongeant graduellement l'intervalle entre les stimuli vert et rouge. Il est évident que si l'intervalle est suffisamment long vous pourrez dire quelle est la différence entre *percevoir* le mouvement et l'inférer. (C'est une nuit sombre et orageuse ; pendant un premier éclair, vous me voyez sur votre gauche ; deux secondes plus tard, il y a un autre éclair et vous me voyez sur votre droite. Vous inférez que je dois avoir bougé, et vous pouvez certainement dire que vous êtes seulement en train d'inférer le mouvement en cette occasion, et pas en train de me *voir* bouger.) Au fur et à mesure que l'expérimentateur allongera les stimuli, il arrivera un moment où vous commencerez à faire cette discrimination. Vous direz par exemple :

« Cette fois-ci, le point rouge ne semble pas avoir bougé, mais après avoir vu le point vert, il m'est venu l'idée que le point rouge a bougé et a changé de couleur. »

En fait, il existe une gamme intermédiaire d'intervalles pour lesquels la phénoménologie est assez paradoxale : vous voyez les points rouges comme deux sources lumineuses stationnaires *et* comme une seule chose qui bouge ! On peut aisément déceler cette sorte de mouvement apparent dans le mouvement plus rapide et plus doux que l'on peut observer au cinéma et à la

télévision, mais notre capacité à faire accomplir cette discrimination n'a pas d'incidence sur le conflit entre le théoricien stalinien et le théoricien orwellien. Ils sont d'accord pour dire que vous faites cette discrimination dans les bonnes conditions. Ce sur quoi ils sont en désaccord, c'est la manière de décrire les cas de mouvement apparent que l'on ne peut *pas* distinguer du mouvement réel, les cas dans lesquels vous percevez réellement le mouvement illusoire. Pour aller vite, ces cas est-ce dans votre mémoire qui vous joue des tours ou bien vos yeux ?

Mais même si vous, le sujet, ne pouvez dire si le phénomène est stalinien ou orwellien, les savants ne pourraient-ils pas – en tant qu'observateurs extérieurs – trouver quelque chose dans votre cerveau qui montre quelle hypothèse est la bonne ? Certains voudront exclure d'emblée cette hypothèse comme *inconcevable*. « Essayez seulement d'imaginer *quelqu'un d'autre* faisant ce dont vous êtes conscient ! Impossible ! » Mais est-ce réellement impossible ? Voyons cela de plus près. Supposez que les savants disposent d'informations vraiment précises (à partir de différentes technologies au scanner) sur le temps exact d'arrivée ou de « création », de tout véhicule de contenu, où que ce soit dans votre système nerveux. Cela leur donnerait le *tout premier* moment de réaction de votre part – conscient ou pas – à n'importe quel contenu (si l'on exclut toute précognition miraculeuse). Mais le temps *effectif* où vous *êtes devenu conscient* de ce contenu (si vous en êtes vraiment devenu conscient) pourrait se situer ultérieurement. Il vous faudrait devenir conscient de ce contenu assez vite pour expliquer le fait que vous incorporiez ce contenu dans un acte de langage ultérieur de souvenir – en admettant que par définition toute entité présente dans votre monde hétérophénoménologique est bien une entité de votre conscience. Cela donnera le moment *ultime* où le contenu est « devenu conscient ». Mais, comme nous l'avons vu, si cela nous laisse en présence d'une durée d'environ plusieurs centaines de millisecondes pendant laquelle la conscience d'une certaine entité doit se produire, et s'il y a plusieurs entités différentes qui doivent survenir pendant cette séquence (le point rouge et le point vert, la femme aux cheveux longs avec ou sans lunettes), il n'y a pas moyen d'utiliser vos *comptes rendus* pour ordonner les événements représentationnels dans la conscience.

Vos comptes rendus verbaux rétrospectifs doivent être neutres relativement à deux possibilités présumées. Mais les savants ne

pourraient-ils pas trouver d'autres données à utiliser ? Ils le pourraient s'il y avait une bonne raison de soutenir qu'un certain type de comportement non verbal (manifeste ou interne) serait un bon indicateur de la conscience. Mais c'est précisément là que les justifications nous manquent. Les deux sortes de théoriciens admettent qu'il n'y a pas de réaction comportementale à un contenu qui ne *pourrait pas* être seulement une réaction inconsciente – sauf pour des verbalisations ultérieures. Selon le modèle stalinien, il existe une action inconsciente de pousser le bouton (pourquoi pas ?). Les deux théoriciens admettent aussi qu'il pourrait exister une expérience consciente qui ne produise pas d'effets comportementaux. Selon le modèle orwellien, il existe une conscience momentanée d'un point rouge stationnaire qui ne laisse pas de trace dans les réactions ultérieures (pourquoi pas ?).

Les deux modèles parviennent sans problème à rendre compte de *toutes* les données – pas seulement des données dont nous disposons déjà, mais aussi de celles que nous imaginons pouvoir obtenir dans le futur. Ils rendent tous deux compte des verbalisations. Selon une des théories, nous nous trompons en toute innocence ; d'après l'autre, il existe des comptes rendus exacts d'erreurs vécues comme telles. De plus, comme nous pouvons le supposer, ces deux sortes de théoriciens ont *exactement* la même théorie de ce qui se passe dans votre cerveau. Ils s'accordent exactement sur le lieu et le moment où, dans le cerveau, le contenu erroné entre dans les circuits causaux ; ils sont seulement en désaccord sur la question de savoir si cette localisation doit être définie comme antérieure à l'expérience ou comme postérieure à elle. Ils donnent la même analyse des effets non verbaux, à une petite différence près. Les uns disent qu'ils sont les résultats de contenus discriminés inconsciemment, et les autres qu'ils sont les résultats de contenus consciemment discriminés mais oubliés par la suite. Finalement, ces théories rendent compte toutes deux des données subjectives – tout ce qui peut être obtenu du point de vue de la première personne – parce qu'elles s'accordent sur ce que peuvent « ressentir » les sujets. Les sujets devraient être incapables de dire quelle est la différence entre des expériences faussées au départ et des expériences immédiatement mal mémorisées.

Par conséquent, malgré les apparences, seules des différences purement verbales séparent ces deux théories (Reingold et Merikle,

en 1990, font un diagnostic similaire). Elles racontent exactement la même chose, sauf en ce qui concerne l'endroit où elles placent la Frontière mythique, un point dans le temps (et par conséquent un lieu dans l'espace) qu'on ne peut localiser de manière fine avec l'aide des sujets. Sa localisation est également neutre en ce qui concerne les autres traits de leurs théories. C'est une différence qui ne fait aucune différence.

Le monde contemporain nous offre une analogie. Dans l'univers de l'édition, on distingue traditionnellement la préparation des prépublications et la correction des *errata* postérieurement à la publication. Mais dans le monde académique d'aujourd'hui les choses vont beaucoup plus vite du fait des publications électroniques. Avec l'avènement du traitement de texte, des publications assistées par ordinateur et du courrier électronique, il arrive souvent que plusieurs versions différentes d'un article se trouvent simultanément en circulation et que l'auteur révisé immédiatement son texte en réponse à des commentaires reçus par courrier électronique. Fixer un moment de publication et appeler ainsi l'une des versions d'un article le texte *canonique*, le texte à enregistrer, celui qu'on citera dans une bibliographie devient une affaire assez arbitraire. Il arrive souvent que la plupart des lecteurs visés, ceux qui comptent, lisent seulement une version antérieure tandis que la version « publiée » est archivée et inerte. Si on cherche à produire des effets importants, la plupart sinon tous sont répartis sur de nombreuses versions ; on n'attend pas la publication. Il fut un temps où les choses étaient différentes et où, virtuellement, tous les effets d'un article important arrivaient *après* la parution en revue et *en raison de* cette parution. À présent que la question de savoir quelle est la version canonique publiée n'est plus importante fonctionnellement – si tant est que cela ait encore une importance quelconque –, il nous faut décider arbitrairement quelle version considérer comme le texte publié. Il n'existe pas de point culminant ou de point décisif dans le chemin conduisant des versions primitives aux archives.

De même – et c'est là une conséquence fondamentale du modèle des Versions Multiples – si l'on veut considérer un certain moment au cours du traitement de l'information dans le cerveau comme celui de la conscience, ce choix doit être arbitraire. On peut toujours « tracer une frontière » dans le flux de la conscience au sein du cerveau, mais il n'y a pas de différence fonctionnelle qui puisse justifier qu'on décrète toutes les étapes antérieures et révi-

sions comme des ajustements inconscients ou préconscients, et toutes les modifications ultérieures du contenu (telles qu'elles apparaissent dans les souvenirs) comme des contaminations post-expérientielles de mémoire. La distinction s'évanouit dans des secteurs proches.

Le théâtre de la conscience revisité

« La technique de l'astronome :
Si tu ne l'écris pas,
cela ne s'est jamais passé. »

Clifford Stoll, *The Cuckoo's Egg*, 1989.

Comme tout manuel de magie vous le dira, les meilleurs tours sont déjà réalisés avant que l'auditoire ne pense qu'ils viennent de commencer. Au point où nous en sommes, vous pourriez penser que je viens précisément de vous en jouer un bon. J'ai soutenu qu'en raison de l'étalement spatio-temporel du point de vue de l'observateur dans le cerveau, aucune de nos données ne nous permet de départager les théories orwelliennes des théories staliniennes de l'expérience consciente, et par conséquent qu'*il n'y a pas de différence*. C'est là une sorte d'opérationnalisme ou de vérificationnisme, qui laisse ouverte la possibilité qu'il existe des faits décisifs que la science n'est pas capable d'atteindre, même quand elle inclut l'hétérophénoménologie. Qui plus est, il semble réellement tout à fait évident qu'il *existe* de tels faits bruts – autrement dit que notre expérience consciente immédiate consiste en de tels faits !

J'admets que cela semble être tout à fait évident. Si ce n'était pas le cas, je n'aurais pas eu besoin de faire tous ces efforts dans ce chapitre pour montrer que ce qui est si évident est en fait faux. Il semble que j'aie laissé quelque chose de côté. Je l'ai fait intentionnellement : ce quelque chose, c'est le Théâtre Cartésien de la Conscience. Sous couvert d'*antidualisme* (« Débarrassons-nous de cette substance étrange et fantomatique »), vous soupçonnez peut-être que j'ai fait s'évaporer (très littéralement) un argument que Descartes avait raison de mettre en avant : il existe un lieu fonctionnel d'une espèce quelconque où les entités phénoménologiques sont... *projetées*.

Il est temps de faire justice de ce soupçon. Nelson Goodman

pose cette question quand il dit de l'expérience du phénomène phi pour les couleurs qu'il « semble nous laisser le choix entre une théorie de la construction rétrospective et une croyance en la voyance » (Goodman, 1978, p. 83). Nous devons rejeter la voyance. Par conséquent, qu'est-ce qu'une « construction rétrospective » ?

« Que l'on considère que la perception du premier flash lumineux soit retardée ou préservée ou ressouvenue, j'appelle cela une théorie de la construction rétrospective – la théorie selon laquelle la construction que l'on perçoit entre les deux flashes ne s'accomplit pas plus tôt que la seconde. »

Au début, Goodman semble hésiter entre une théorie stalinienne (la perception du premier flash est retardée), et une théorie orwellienne (la perception du premier flash est préservée ou gardée en mémoire), mais plus important encore, le révisionniste qu'il postule (qu'il soit orwellien ou stalinien) ne se contente pas d'ajuster des jugements ; il *construit* des matériaux destinés à *combler* les lacunes :

« Chacun des endroits le long d'un chemin situé entre les deux flashes est rempli avec l'une des couleurs qui s'allument plutôt qu'avec des couleurs intermédiaires successives » (p. 85).

Goodman ne voit pas la possibilité suivante : que le cerveau n'ait pas en fait à s'occuper de « remplir » quoi que ce soit avec une « construction » – car il n'y a personne pour la regarder. Comme le modèle des Versions Multiples le rend explicite, une fois qu'une discrimination a été faite, il est inutile qu'elle se répète ; le cerveau se contente d'ajuster la conclusion qui est tirée, faisant la nouvelle interprétation de l'information disponible pour la modulation du comportement qui suit.

Goodman a en vue la théorie, qu'il attribue à Van der Waals et Roelofs (1930), selon laquelle « le mouvement intermédiaire est produit rétrospectivement, construit seulement après que le second flash ne survienne, et *projeté en arrière dans le temps* (mes italiques) » (p. 73-74). Cela suggère une théorie stalinienne comportant une variante incroyable : on fabrique un film final et on le passe dans un projecteur magique dont le rayon voyage en quelque sorte à rebours dans le temps sur l'écran de l'esprit. Ce n'est peut-être pas exactement ce que Van der Waals et Roelofs avaient en tête quand ils ont proposé les « reconstructions rétrospectives », mais c'est sans doute ce qui a conduit Kolers (1972,

p.184) à rejeter leur hypothèse : selon Kolers, en effet, toute construction a lieu « en temps réel ». Et pourtant, pourquoi le cerveau se préoccuperait-il de « produire » le « mouvement intermédiaire » en temps réel ? Pourquoi le cerveau ne se contenterait-il pas simplement de *conclure qu'il existait un mouvement intermédiaire*, et pourquoi n'insérerait-il pas cette conclusion rétrospective dans le flux de son traitement ? N'est-ce pas suffisant ?

Halte ! C'est là que le tour de passe-passe (s'il y en a un) doit se produire. Du point de vue de la troisième personne, j'ai postulé un sujet, le sujet hétérophénoménologique, qui serait une sorte de « à qui de droit » fictionnel auquel, bien entendu, les observateurs extérieurs attribueraient correctement la croyance qu'un certain mouvement intermédiaire a été éprouvé. C'est comme cela que les choses apparaîtront à *ce* sujet (qui n'est qu'une fiction produite par le théoricien). Mais n'existe-t-il pas aussi un sujet *réel* au bénéfice duquel le cerveau doit monter un spectacle, remplissant toutes les taches aveugles ? C'est ce que Goodman semble supposer quand il dit que le cerveau remplit les points situés sur le chemin. Qui bénéficie de ce dessin animé ? L'auditoire du Théâtre Cartésien. Mais *comme il n'y a pas de théâtre, il n'y a pas d'auditoire non plus*.

Le modèle des Versions Multiples s'accorde avec les thèses de Goodman : rétrospectivement, le cerveau ne se préoccupe pas de « construire » des représentations qui auraient à charge de « remplir » les lacunes. Ce serait une perte de temps et (oserais-je le dire ?) de *peinture*. Le jugement est *déjà là*, et le cerveau peut ainsi se consacrer à d'autres tâches ⁶.

La « projection à rebours dans le temps » dont parle Goodman est une expression ambiguë. Elle peut avoir un sens modeste et défendable : à savoir qu'une *référence à un moment passé* est incluse dans le contenu. Selon cette interprétation, ce serait comme dire « ce roman nous emmène dans la Rome ancienne ». Personne ne songerait à interpréter cet énoncé comme une extravagance métaphysique : le roman n'agit pas comme une sorte de machine à remonter le temps. C'est l'interprétation qui s'accorde avec les autres thèses avancées par Goodman. Mais Kolers a, semble-t-il, cru qu'elle avait un sens métaphysiquement radical : une chose aurait été réellement projetée d'un temps donné à un autre temps.

Comme nous le verrons dans le chapitre suivant, la confusion provoquée par cette interprétation radicale de la notion de « projection » a infecté l'interprétation des autres phénomènes. C'est

la même métaphysique curieuse qui hantait nos conceptions de la représentation de l'espace. À l'époque de Descartes, Thomas Hobbes semble avoir pensé qu'après que la lumière a touché les yeux et produit une sorte de mouvement dans le cerveau, cela conduit quelque chose à *se lier* dans le monde.

« La cause de la sensation est le corps extérieur, ou objet, qui presse l'organe propre à chaque sensation, soit immédiatement, comme dans le goût et le toucher, soit médiatement, comme dans la vue, l'ouïe, et l'odorat ; cette pression, propagée vers l'intérieur, par l'intermédiaire des nerfs ainsi que des autres fibres ou membranes du corps, jusqu'au cerveau et au cœur, cause là une résistance, une contre-pression, un effort du cœur pour se délivrer ; cet effort, étant dirigé vers *l'extérieur*, semble être quelque réalité située au-dehors » (*Léviathan*, partie I, ch. 1, « De la sensation », trad. Tricaud, p. 12-13).

Après tout, pensait-il, c'est là que nous voyons les couleurs — sur les surfaces des objets⁷ ! Dans le même esprit, on pourrait supposer que quand vous vous cognez le doigt de pied, cela cause des signaux ascendants aux « centres de douleur » du cerveau, qui « projettent » ensuite la douleur *vers le doigt de pied là où elle doit se trouver*. Après tout, c'est là qu'on éprouve la douleur.

Même récemment, dans les années cinquante, cette idée était prise suffisamment au sérieux pour qu'un psychologue britannique, J.R. Smythies, écrive un article pour la démolir⁸. La projection dont nous parlons quand nous avons affaire à ces phénomènes n'implique pas que nous produisons un effet quelconque dans l'espace physique, et je pense que personne ne pense plus que c'est le cas. Mais les neurophysiologues, les psychologues et les théoriciens de l'acoustique qui construisent des systèmes stéréophoniques parlent souvent de ce genre de projection. Nous pourrions nous demander si ce n'est pas à une transmission physique d'un lieu (ou d'un temps) à un autre qu'ils pensent. Qu'est-ce que cela veut dire ? Prenons un cas simple :

« Du fait que l'on a placé les haut-parleurs stéréo et modulé le volume de leurs outputs respectifs, l'auditeur *projette* le son résultant du soprano vers un point se trouvant à mi-chemin entre les deux haut-parleurs. »

Qu'est-ce que cela veut dire ? Il nous faut analyser ce point avec précaution. Si les haut-parleurs sont en train de gueuler dans une pièce vide, il n'y a pas de projection du tout. S'il y a

un auditeur présent (un observateur doté d'une bonne oreille, et d'un bon cerveau), la « projection » se produit, mais cela ne veut *pas* dire que quelque chose soit émis par l'auditeur à un point situé à mi-chemin entre les deux haut-parleurs. Aucune propriété de ce point ou d'un point voisin ne se trouve changée du fait de la présence de l'auditeur. Bref, c'est ce que nous voulons dire quand nous disons que Smythies avait raison : il n'y a pas de projection dans l'espace de propriétés visuelles ou auditives. Que se passe-t-il alors ? Il *semble* à l'observateur que le son du soprano vient de ce point. Mais qu'implique cet *apparaître à l'observateur* ? Si nous répondons que cela implique « la projection par l'observateur du son à ce point dans l'espace » nous revenons de toute évidence à notre point de départ, et par conséquent les gens sont tentés d'introduire quelque chose d'autre et de dire par exemple : « L'observateur projette le son dans un espace *phénoménal*. » On n'a pas fait un grand progrès, semble-t-il. Nous avons nié que la projection ait lieu dans l'espace physique, mais avons relocalisé la projection dans l'espace phénoménal.

Mais qu'est-ce que l'espace phénoménal ? Est-ce un espace physique à l'intérieur du cerveau ? Est-ce l'espace scénique dans un théâtre de la conscience situé dans le cerveau ? Pas au sens littéral. Alors dans un sens métaphorique ? Au chapitre précédent, nous avons vu que nous pouvions envisager de tels espaces métaphoriques, dans le cas des « images mentales » manipulées par Shakey. En un sens strict bien que métaphorique, Shakey dessinait des formes dans l'espace, était attentif à des points particuliers de l'espace, et fondait ses conclusions sur ce qu'il découvrait à propos de points particuliers dans cet espace. Mais l'espace en question était un espace *logique*. Comme l'espace du Londres de Sherlock Holmes, c'était l'espace d'un monde fictionnel certes, mais systématiquement ancré dans des événements physiques réels ayant lieu dans l'espace ordinaire comme le « cerveau » de Shakey. Si nous considérions les énoncés de Shakey comme des expressions de ses « croyances » nous pourrions dire que c'était un espace auquel Shakey *croyait*. Mais pour autant il n'en devenait pas réel, pas plus que la croyance de quelqu'un en Phénhomme ne rendrait Phénhomme réel. Tous deux ne sont que des objets intentionnels⁹.

Nous donnons donc bien à la notion d'espace phénoménal le sens d'un espace logique. C'est un espace vers lequel ou dans lequel rien n'est littéralement projeté ; ses propriétés sont sim-

plement constituées par les croyances du sujet (hétérophénoménologiques). Quand nous disons que l'auditeur projette le son vers un point dans cet espace, nous voulons *seulement* dire qu'il lui semble que c'est de là que vient le son. Cela n'est-il pas suffisant ? Ou bien ne sommes-nous pas en train de négliger une doctrine « réaliste » de l'espace phénoménologique dans lequel *l'apparence réelle* de l'objet pourrait être projetée ?

Aujourd'hui, nous sommes familiers de la distinction entre la localisation spatiale dans le cerveau du véhicule de l'expérience et la localisation dans l'« espace de l'expérience » des entités dont nous faisons l'expérience. En bref, nous distinguons le représentant du représenté, le véhicule du contenu. Nous sommes devenus assez sophistiqués pour reconnaître que les produits de la perception visuelle ne sont pas, littéralement, des images dans la tête même si *ce qu'elles représentent* est quelque chose que les images représentent bien : la disposition dans l'espace de diverses propriétés visibles. Nous devrions formuler la même distinction pour le temps : *le moment où*, dans le cerveau, une expérience se produit doit être distingué de ce qui semble se produire. En fait, comme le psycholinguiste Ray Jackendoff l'a suggéré, ce qu'il nous faut ici comprendre est en fait une extension banale de nos idées courantes quant à la perception de l'espace. La représentation de l'espace dans le cerveau n'utilise pas toujours l'espace-dans-le-cerveau pour représenter l'espace, et la représentation du temps dans le cerveau n'utilise pas toujours le-temps-dans-le-cerveau. Le projecteur de films que postule l'interprétation radicale de la « projection en arrière dans le temps » dont parlait Goodman est tout aussi infondée que les projecteurs de diapositives spatiales que Smythies refusait de supposer dans le cerveau.

Pourquoi les gens éprouvent-ils le besoin de postuler l'existence de ce projecteur d'apparences ? Pourquoi sont-ils enclins à penser qu'il ne suffit pas pour les salles de rédaction du cerveau de simplement insérer des contenus dans le flux qui mène à la modulation du comportement et de la mémoire ? Peut-être parce qu'ils veulent préserver pour la conscience la distinction entre apparence et réalité. Ils veulent résister à l'opérationnalisme diabolique qui dit que ce qui s'est produit (dans la conscience) est simplement ce que vous vous rappelez s'être produit. Le modèle des Versions Multiples fait de « l'écriture dans la mémoire » le critère même de la conscience ; c'est *ce en quoi consiste* le fait

que le « donné » soit « pris » comme ceci plutôt que comme cela. L'expérience consciente des effets de divers véhicules du contenu sur des actions ultérieures (et par conséquent aussi sur la mémoire) n'a pas de réalité. Cela ressemble de façon inquiétante à un effrayant opérationnalisme, et peut-être le Théâtre Cartésien de la conscience est-il chéri en cachette comme le lieu où tout ce qui se passe « dans la conscience » arrive *réellement*, que ce soit ou non correctement mémorisé. Supposez que quelque chose se soit produit en ma présence mais n'ait laissé de trace en moi que « pour un millionième de seconde », comme dans l'épigramme d'Ariel Dorfman. Qu'est-ce que cela peut bien vouloir dire que j'aie été, aussi brièvement et sans effet sur moi que possible, conscient de cette chose ? S'il y avait un Théâtre Cartésien privilégié quelque part, cela pourrait au moins signifier que *c'était là un joli film* même si personne ne se souvient de l'avoir vu. (Nous y voilà !)

Le Théâtre Cartésien est une image réconfortante sans doute parce qu'il préserve la distinction apparence-réalité au cœur de la subjectivité humaine, mais il est sans assise scientifique. De même, il est métaphysiquement douteux, parce qu'il crée une catégorie bizarre, celle de choses objectivement subjectives, de choses qui, réellement et objectivement, vous apparaissent, même si elles ne vous apparaissent pas comme étant telles ! (Smullyan, 1981.) Certains penseurs s'opposent si violemment au « vérificationnisme » et à l'« opérationnalisme » qu'ils veulent les rejeter même dans le domaine où ceux-ci ont un sens : le domaine de la subjectivité. Ce que Clifford Stoll appelle la technique de l'astronome est un commentaire sarcastique sur les égarements de la mémoire et les critères des confirmations scientifiques, mais il devient une vérité littérale quand on l'applique à ce qui est « écrit » dans la mémoire. Nous pourrions ranger, par conséquent, le modèle des Versions Multiples dans la catégorie de *l'opérationnalisme de la première personne*, parce qu'il nie tout de go et par principe la possibilité d'avoir la conscience d'un stimulus en l'absence de la croyance du sujet en cette conscience¹⁰.

Quand on s'oppose à cet opérationnalisme, on fait appel, comme d'habitude, à des faits possibles situés au-delà de la sphère à laquelle l'opérationnaliste peut avoir accès pour confirmer ses hypothèses, mais dans le cas qui nous occupe, l'opérationnaliste est le sujet lui-même, et par conséquent on peut retourner l'objection : « C'est précisément parce que vous ne pouvez pas dire,

d'après vos critères préférés, si vous étiez conscient ou pas de la chose x, que cela ne veut pas dire que vous ne l'étiez pas. Peut-être étiez-vous conscient de x, mais ne pouviez pas le vérifier ! » Quelqu'un a-t-il, à la réflexion, envie de dire quelque chose de ce genre ? Des faits putatifs concernant la conscience qui nagent, hors de la portée des observateurs « extérieurs » aussi bien que des observateurs « intérieurs », sont des faits bien étranges.

C'est une idée qui a la vie dure. Songez à l'expression « j'ai jugé que les choses étaient ainsi parce que c'est comme cela qu'elles m'ont semblé être ». Ici, on nous encourage à penser à deux états ou événements distincts : le fait d'apparaître-d'une-certaine-façon et le fait qui s'ensuit de juger-que-les-choses-sont-ainsi. Le problème, peut-on penser, avec le modèle des Versions Multiples pour la couleur phi, par exemple, est que même s'il inclut le phénomène du sujet qui juge qu'il y a eu un mouvement intermédiaire, il n'inclut pas – il nie même explicitement son existence – un quelconque événement que l'on pourrait appeler le mouvement-qui-sembler-être-intermédiaire, sur lequel ce jugement est « fondé ». Il doit exister « des données présentées » quelque part, ne serait-ce que dans un faux procès stalinien, de manière à ce que le jugement soit fondé sur ces données.

Il existe des gens qui présument que cette intuition est justifiée par la phénoménologie. Ils ont l'impression qu'ils s'observent effectivement eux-mêmes en train de juger des choses comme étant telles ou telles *en vertu du fait* que ces choses leur semblent être telles. Personne n'a jamais observé quoi que ce soit de ce genre « dans sa phénoménologie », parce qu'un tel fait de causalité serait inobservable (comme Hume l'a noté il y a fort longtemps¹¹).

Demandez à un sujet soumis à l'expérience de la couleur phi s'il juge que le point rouge s'est mû vers la droite et a changé de couleur parce qu'il lui a semblé qu'il le faisait, ou bien s'il lui semble qu'il a bougé parce qu'il en a jugé ainsi. Supposez que le sujet donne une réponse « sophistiquée » :

« Je sais qu'il n'y avait pas *en fait* de point en mouvement dans le monde – ce n'est que du mouvement apparent, après tout – mais je sais aussi que le point *semblait* se mouvoir, en sorte qu'en plus de mon jugement que le point semblait se mouvoir, il y a l'événement *au sujet duquel* porte mon jugement : l'apparence-de-mouvement du point. Il n'y avait pas de véritable mouvement, et donc

il devait y avoir une véritable apparence-de-mouvement pour mon jugement. »

Peut-être le Théâtre Cartésien est-il populaire parce que c'est le lieu où les apparences peuvent se produire en plus des jugements. Mais l'argument sophistiqué qui vient d'être présenté est fallacieux. Postuler une « apparence réelle » *en plus* du jugement ou de la façon de prendre les choses exprimées par le compte rendu du sujet revient à multiplier les entités au-delà de ce qui est nécessaire. Pire, cela revient à multiplier les entités au-delà de ce qui est possible ; la forme de présentation interne dans laquelle des apparences véritables ont lieu est une échappatoire métaphysique désespérée, une façon d'essayer d'avoir le beurre et l'argent du beurre, en particulier parce que ceux qui sont enclins à penser ainsi sont prompts à insister sur le fait que cette représentation interne ne se produit pas dans une sorte d'espace mystérieux, dualiste, rempli d'un éther de fantômes cartésiens. Quand vous rejetez le dualisme cartésien, vous devez en fait rejeter la pièce qu'on aurait représentée sur la scène du Théâtre Cartésien, et l'auditoire lui-même, car on ne trouve ni pièce ni auditoire dans le cerveau, alors même que le cerveau est le seul endroit où l'on puisse les trouver.

Le modèle des Versions Multiples en action

Passons en revue le modèle des Versions Multiples, en l'étendant un peu, et en considérant un peu plus en détail la situation dans le cerveau qui le justifie. Pour simplifier, je me concentrerai sur ce qui se passe dans le cerveau pendant l'expérience visuelle. Nous étendrons plus tard l'analyse à d'autres phénomènes.

Les stimuli visuels provoquent des chaînes d'événements dans le cortex qui produisent graduellement des discriminations de plus en plus spécifiques. À différents temps et dans différents lieux, diverses « décisions » et différents « jugements » se produisent ; plus précisément, des parties du cerveau viennent à se trouver dans de nouveaux états qui discriminent différents traits, par exemple d'abord le simple point de départ d'un stimulus, puis la localisation, puis la forme, puis plus tard la couleur (dans un chemin différent), plus tard encore le mouvement (apparent), et finalement la reconnaissance des objets. Ces états discriminatoires localisés transmettent des effets à d'autres lieux, contri-

buant à des discriminations ultérieures, et ainsi de suite (Van Essen, 1979 ; Allman, Meizin et Mc Guinness, 1985 ; Livingstone and Hubel, 1987 ; Zeki and Shipp 1988). La question naturelle mais naïve qu'on peut poser est la suivante : où donc toutes ces choses se rejoignent-elles ? La réponse est : nulle part. Certains de ces états distribués porteurs de contenus s'évanouiront rapidement, sans laisser de traces. D'autres laisseront des traces, sur des comptes rendus verbaux ultérieurs d'expérience et de mémoire, sur la « capacité sémantique » et sur d'autres sortes de dispositifs perceptifs, sur les états émotionnels, les tendances comportementales, et ainsi de suite. Certains de ces effets, par exemple des influences sur des verbalisations ultérieures, sont au moins des symptômes de l'activité consciente. Mais il n'y a pas de lieu dans l'esprit à travers lequel tous ces processus causaux doivent passer pour déposer leur contenu « dans la conscience ».

À partir du moment où une discrimination de ce genre s'est accomplie, elle devient disponible pour produire un certain comportement, par exemple pousser un bouton (ou produire un sourire, ou un commentaire) ou pour moduler un état informationnel interne quelconque. Par exemple, la discrimination de l'image d'un chien peut créer un « dispositif perceptif » – qui rend plus rapide l'identification visuelle des chiens (ou d'autres animaux) dans d'autres images – ou bien elle peut activer un certain domaine sémantique, en rendant plus probable que vous lisiez le mot « aboie » comme désignant un son, et non comme une situation dans laquelle on a peur. Comme nous l'avons noté, ce processus distribué se déroule pendant des milliers de millisecondes, pendant lesquelles différentes additions, incorporations, amendements et réécritures de contenus peuvent survenir, dans des ordres divers. Ceux-ci finissent par produire, au bout du compte, quelque chose qui ressemble *plutôt* à un flux narratif ou à un récit suivi, qu'on peut considérer comme sujet à un processus d'édition constant par des nombreux processus distribués alentour dans le cerveau et se poursuivant dans le futur. Les contenus surgissent, sont révisés, contribuent à l'interprétation d'autres contenus ou à la modulation du comportement (verbal ou autre) et, dans ce processus, laissent leur trace dans la mémoire, qui ensuite peut finalement s'estomper ou être incorporée dans ou être réécrite par d'autres contenus, en totalité ou en partie. Cet amas de contenus est plutôt comme un récit en raison de sa multiplicité ; à tout point du temps, il existe des

versions multiples des fragments narratifs à différentes étapes du travail d'édition et dans différents endroits du cerveau. Alors que certains des contenus dans ces diverses versions auront une contribution très brève et disparaîtront sans autres effets – et que certains n'auront aucune contribution du tout – d'autres persisteront à jouer toute une variété de rôles dans la modulation ultérieure de l'état interne et du comportement, et quelques-uns persisteront jusqu'à faire connaître leur présence à travers des dépêches de presse émises sous forme de comportement verbal.

Intervenir dans ce flux à différents intervalles produit différents effets, précipitant différents récits. Ce *sont* bien cette fois des récits, des versions isolées d'une portion du « flux de la conscience ». Si l'on retarde cette intervention trop longtemps, le résultat risque d'être qu'il n'y aura aucun récit. Si l'on intervient « trop tôt », on peut saisir des données au moment où une discrimination particulière a été produite dans le flux, mais on interrompt alors la progression normale du flux.

Existe-t-il un moment « optimal d'intervention » ? Si l'on forme l'hypothèse plausible d'après laquelle, après un certain temps, les récits se dégradent assez rapidement à la fois parce que les détails s'estompent et en raison des embellissements qu'on leur fait subir (ce que j'aurais dû dire à la soirée tend à devenir ce que j'ai effectivement dit à la soirée), on peut justifier une intervention aussi rapide que possible après la suite de stimuli pertinents. Mais on veut aussi éviter d'interférer dans le phénomène en intervenant trop tôt. Étant donné que la perception se transforme imperceptiblement en mémoire et que l'interprétation « immédiate » se transforme imperceptiblement en reconstruction rationnelle, il n'existe pas de point culminant valable dans tous les contextes vers lequel nous pourrions diriger nos interventions.

Ce dont nous sommes conscients pendant une durée temporelle particulière ne se définit pas indépendamment des interventions que nous utilisons pour précipiter une narration au sujet de cette période. Étant donné que ces récits sont en révision perpétuelle, il n'existe pas de récit unique qui doive compter comme version canonique, comme « première édition » dans laquelle seraient couchés, une fois pour toutes, les événements qui se sont produits dans le flux de la conscience du sujet et par rapport à laquelle toutes les déviations seraient des corruptions du texte. Mais tout récit (ou fragment de récit)

qui n'est pas précipité fournit une « ligne temporelle », une séquence subjective d'événements survenant dans le cerveau de l'observateur. Comme nous le verrons, ces deux lignes temporelles *peuvent* ne pas se superposer par un enregistrement orthogonal (bien alignées). Même si c'était le cas, la (fausse) discrimination de *rouge-devenant-vert* est survenue dans le cerveau *après* la discrimination de *point rouge*, *puis rouge-devenant-vert*, et *finalement point vert*. Par conséquent, pendant l'étalement temporel des points de vue du sujet, il peut exister d'autres différences qui peuvent induire des brouillages.

Cet échec de nos enregistrements n'a rien de métaphysiquement extravagant ou de problématique ¹². Il n'est pas plus mystérieux ou contraire à la causalité que le fait de comprendre que les scènes individuelles dans les films sont souvent envoyées en désordre, en sorte que quand vous lisez la phrase « Daniel arriva à la soirée après Joëlle, mais Monique arriva avant eux », vous apprenez l'arrivée de Daniel avant d'apprendre l'arrivée antérieure de Monique. L'espace et le temps de la représentation sont un cadre de référence ; l'espace et le temps de ce que représente la représentation en constituent un autre. Mais ce fait métaphysiquement bénin est cependant au fondement d'une autre catégorie métaphysique fondamentale : quand une certaine portion du monde vient ainsi à composer un entrelacs de récits, cette portion du monde est un observateur. C'est ce en quoi consiste le fait d'être un observateur dans le monde, un effet que cela fait.

Ce n'est là qu'une esquisse rapide du modèle que je propose. Il faut encore voir plus clairement en quoi il diffère du Modèle

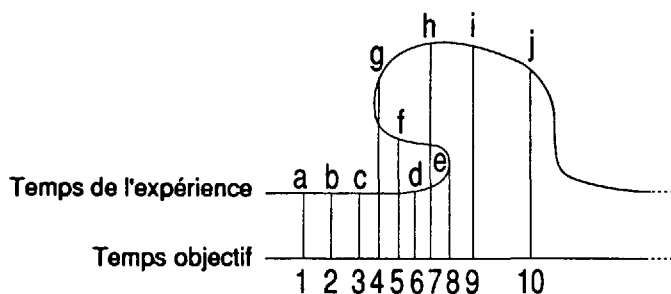


Figure 4.12

du Théâtre Cartésien, en montrant comment il rend compte de certains phénomènes particuliers. Dans le chapitre suivant, nous ferons travailler ce modèle sur certaines questions difficiles. Envisageons d'abord quelques exemples courants, souvent discutés par les philosophes.

Vous avez probablement déjà fait l'expérience suivante : conduire pendant des kilomètres tout en conversant avec votre passager (ou tout en soliloquant), et découvrir ensuite que vous n'avez aucun souvenir de la route, du trafic ou de vos activités de conducteur. Bien des théoriciens (y compris moi-même, je le reconnais, Dennett 1969, p. 116 *sq*) y ont vu leur exemple favori de « perception et d'action intelligente inconscientes ». Mais étiez-vous réellement inconscient de toutes ces autos qui passaient, des feux rouges, des virages au moment où vous avez eu ces activités ? Vous faisiez attention à d'autres choses. Mais *si l'on avait attiré votre attention* sur ce que vous veniez *juste* de voir à différents moments du voyage, vous auriez au moins rapporté quelques détails sommaires. Le phénomène de la « conduite inconsciente » apparaît bien plus comme un cas de conscience flottante avec de rapides pertes de mémoire que comme un cas d'action inconsciente.

Êtes-vous constamment conscient du tic-tac de l'horloge ? S'il s'arrête soudainement, vous vous en apercevrez, et vous verrez tout de suite qu'il s'est arrêté ; les tic-tac dont « vous n'auriez jamais eu conscience » s'ils ne s'étaient pas arrêtés sont à présent clairement dans votre conscience. Un cas encore plus frappant est celui où vous êtes capable de compter, rétrospectivement dans son expérience mémorielle, les coups de l'horloge dont vous avez seulement remarqué qu'elle sonnait après quatre ou cinq coups. Mais comment pourriez-vous aussi clairement vous *souvenir avoir entendu* quelque chose dont vous n'avez pas été conscient au premier chef ? Cette question trahit un présupposé en faveur du modèle cartésien ; quant au flux de la conscience, il n'existe pas de faits fixes indépendants des sondages particuliers.

NOTES

1. Cela rappelle les difficultés que rencontrent les physiciens quand ils sont confrontés à une *singularité*, un point où, précisément en raison de son absence de dimension, différentes grandeurs sont infinies (étant donné leurs définitions). Cela se produit dans le cas des trous noirs, mais affecte aussi l'interprétation d'entités plus familières. Roger Penrose discute la question de savoir comment appliquer les transformations de Lorentz et les équations de Maxwell aux particules. Ces équations font intervenir le champ électromagnétique au *point* précis où la particule chargée se trouve localisée (ce champ est nécessaire à la détermination de la « force » au point considéré). Mais si la particule a une taille finie, quel point choisir ? Faut-il prendre le « centre » de la particule, ou bien ne faut-il pas plutôt considérer (pour le calcul de la « force ») la valeur moyenne du champ aux divers points de la surface de la particule ? ... Peut-être vaut-il mieux considérer la particule comme *ponctuelle*. Mais ce sont alors d'autres problèmes qui se dressent. Car dans ce cas le champ électrique propre à la particule devient *infini* en voisinage immédiat (Penrose, 1989, p. 189-190, trad. fr. p. 204).

2. Ou Matignon, en temps de cohabitation (NdT)

3. Nier que la tête est un quartier général serait pure folie, mais elle ne serait pas sans antécédents. Philippe Pinel a rapporté en 1800 la maladie curieuse d'un homme qui tomba dans un véritable délire produit par les terreurs de la révolution. La perte de sa raison est marquée par une singularité particulière : il croit qu'il a été guillotiné, et que sa tête a été jetée pêle-mêle sur la pile de celles des autres victimes, et que les juges, se repentant trop tard de leur geste cruel, avaient ordonné que les têtes soient prises et recollées sur leurs corps respectifs. Cependant, par l'effet d'une erreur quelconque, ils ont mis sur ses épaules la tête d'un autre infortuné. Cette idée que sa tête a été changée l'occupe nuit et jour... « Regardez mes dents, répétait-il incessamment, elles étaient merveilleuses, et elles sont pourries ! Ma bouche était en bonne santé et *celle-ci* est infectée ! Quelle différence entre ces cheveux et ceux que j'avais avant mon changement de tête ! » *Traité médico-philosophique sur l'aliénation mentale, ou la Manie*, Paris, chez Richard, Caille et Ravier, 1800. (Je remercie Dora Weinberg d'avoir attiré mon attention sur ce cas fascinant.)

4. Un exemple encore plus frappant est une expérience dans laquelle le sujet est trompé par des miroirs en croyant qu'il est en train de regarder sa propre main qui dessine une ligne, alors qu'en fait il est en train de regarder la main du complice de l'expérimentateur. Dans ce cas, « les yeux gagnent » à un tel point que le processus éditorial dans le cerveau est trompé et conclut que la main du sujet est mue de force ; le sujet prétend ressentir la « pression » qui empêche « sa » main de bouger dans le sens où elle est supposée bouger (Nielsen, 1963).

5. L'opérationnalisme est (en gros) la thèse ou la démarche qui s'exprime par la déclaration : « Si vous ne pouvez découvrir une différence, il n'y a pas de différence. » Ou, comme on dit souvent, « si cela couaque comme un canard, et marche comme un canard, c'est un canard ». J'ai reconsidéré les forces et faiblesses de l'opérationnalisme dans Dennett, 1985a.

6. Il y a une région dans le cortex appelée MT, qui répond au mouvement (et au mouvement apparent). Supposez alors qu'il y ait une activité dans MT qui *soit* pour le cerveau le fait de conclure qu'il y avait un mouvement intermédiaire. Dans le Modèle des Versions Multiples, nous n'avons pas besoin de nous demander s'il y a une conclusion antérieure ou postérieure à l'expérience. Autrement dit ce serait une erreur que de se demander si l'activité dans MT est « une réaction à une expérience consciente » (selon l'historien orwellien) par opposition à une « décision de représenter le mouvement » (selon l'éditeur stalinien).

7. En fait, Hobbes avait bien conscience des problèmes posés par cette thèse : « Car si ces couleurs et ces sons étaient dans les corps ou objets qui les causent, on ne pourrait pas les en séparer, comme nous voyons qu'ils le sont par les miroirs, ou par réflexion, dans les échos ; là nous savons que la chose que nous voyons est dans un endroit et sans apparition dans un autre » (*Leviathan*, même chapitre). Mais ce passage se prête à plusieurs interprétations.

8. Smythies (1954). Cet article héroïque montre combien ces sujets étaient difficiles même il y a trente ans. Il prend la peine de critiquer une version de la théorie de la projection énoncée par un manuel, et dans son résumé, il cite avec approbation le rejet de cette même idée par Bertrand Russell : « Quiconque accepte la théorie causale de la perception est obligé de conclure que les percepts sont dans notre tête, parce qu'ils arrivent au terme d'une chaîne causale d'événements physiques conduisant, spatialement, de l'objet au cerveau de celui qui perçoit. Nous ne pouvons pas supposer que, au terme de ce processus, le dernier effet saute soudainement en arrière au point de départ, comme une corde tendue qui se détend » (Russell, 1927).

9. « C'est comme si notre Phénommaniste devenu Phénomménologue venait dans sa confusion à recourir au stratagème désespéré consistant à inventer un espace-dieu, ou un paradis, pour qu'y réside son bien aimé Phénomme, un espace assez *réel* pour satisfaire celui qui croit en lui, mais en même temps suffisamment éloigné et mystérieux pour cacher Phénomme aux yeux du sceptique qu'il y a en lui. L'espace phénoménal est un Paradis en Image Mentale, mais si les images mentales se trouvent en fait être réelles, elles peuvent résider très confortablement dans l'espace physique de nos cerveaux, et s'il se révèle qu'elles ne sont pas réelles, elles peuvent résider, comme le Père Noël, dans l'espace logique de la fiction » (Dennett, 1978 a, p. 186).

10. Le philosophe Jay Rosenberg m'a fait remarquer que Kant voit en cela la marque de la sagesse, quand il dit que, dans l'expérience le *für mich* (le « pour moi ») et le *an sich* (« l'en soi ») sont la même chose.

11. Le philosophe Ned Block m'a raconté son expérience comme sujet dans un test de « latéralisation ». Il regardait droit devant lui vers un point fixé et chaque fois qu'il faisait cela, un mot (ou quelque chose qui n'était pas un mot, comme CHIRPE) s'allumait sur la gauche ou la droite du point fixé. Sa tâche consistait à presser un bouton si le stimulus était un mot. Ses temps de réaction étaient, une fois mesurés, plus longs pour des mots envoyés du côté gauche (et par conséquent entraient dans l'hémisphère gauche en premier), ce qui étayait l'hypothèse qu'il était, comme la plupart des gens, fortement latéralisé pour le langage dans l'hémisphère gauche. Cela n'avait rien de surprenant pour Block. Ce qui l'intéressait était « la phénoménologie : les mots s'allumant sur la gauche semblaient un peu flous, en quelque façon ». Je lui demandai s'il pensait que les mots étaient plus difficiles à identifier. Il admit qu'il n'avait aucun moyen de distinguer ces analyses causales « opposées » de son jugement.

12. Cette façon de voir les choses m'a été suggérée par la lecture de Snyder (1988), bien que sa manière d'envisager les problèmes soit assez différente de la mienne.

Chapitre 5

Le temps et l'expérience

« Je peux certes dire que mes représentations se suivent les unes les autres ; mais cela signifie seulement que nous sommes conscients d'elles dans le temps, c'est-à-dire en conformité avec la forme du sens interne. »

Kant, *Critique de la raison pure*, 1781.

Dans le chapitre précédent, nous avons commencé à voir comment le modèle des Versions Multiples dissout le problème des « projections à rebours dans le temps », mais nous avons laissé de côté certaines difficultés majeures. Dans ce chapitre, nous allons poursuivre cette analyse sur un terrain plus glissant. Nous allons examiner et résoudre plusieurs controverses qui ont surgi chez les psychologues et les chercheurs en neurosciences au sujet de l'explication qu'on peut donner de certaines expériences notoirement problématiques. Je pense qu'il est possible de comprendre le reste du livre sans suivre tous les arguments qui figurent dans ce chapitre, en sorte qu'on peut le passer ou le laisser de côté, mais j'ai essayé de rendre ces sujets accessibles aux débutants. Je vois au moins six raisons d'entreprendre cette analyse technique.

1. Ma présentation sommaire du modèle des Versions Multiples comporte encore quelques obscurités. En voyant le modèle fonctionner, vous aurez une vision plus claire de sa structure.

2. Si vous avez encore des doutes et si vous vous demandez en quoi le modèle des Versions Multiples diffère, en tant que

théorie empirique, du Théâtre Cartésien traditionnel, ces doutes se dissiperont au vu de plusieurs collisions frontales entre les deux modèles.

3. Si vous vous demandez si je ne suis pas en train de tirer sur un mort, il sera rassurant de découvrir quelques experts prisonniers des nœuds dans lesquels ils se sont eux-mêmes emberlificotés parce qu'ils sont matérialistes cartésiens malgré eux.

4. Si vous me soupçonnez d'avoir fondé mon modèle sur un seul phénomène habilement choisi, le phi pour les couleurs de Kolers, vous verrez comment des phénomènes très différents se prêtent aussi au traitement en termes de Versions Multiples.

5. Certaines des expériences célèbres que nous allons examiner ont été montées en épingle par *certain*s spécialistes comme réfutant le type de théorie matérialiste conservatrice que je présente ici. Si par conséquent il y a un défi *scientifique* opposé à mon explication de la conscience, c'est le champ de bataille qui a été choisi par l'opposition.

6. Enfin, les phénomènes en question sont fascinants et méritent qu'on s'en occupe ¹.

Moments flottants et lapins bondissants

Une condition habituellement suffisante, mais pas nécessaire, pour avoir eu l'expérience de quelque chose est un compte rendu (*report*) verbal ultérieur. C'est là le cas autour duquel tournent tous les phénomènes problématiques. Supposez que bien que votre cerveau ait enregistré – répondu à – certains aspects d'un événement, quelque chose intervienne entre cette réponse interne et une occasion ultérieure que vous pouvez avoir de faire un compte rendu verbal. S'il n'y avait pas de temps ni de possibilité pour vous de formuler une réponse initiale explicite, et si des événements intervenaient de manière à empêcher que des réponses explicites ultérieures (verbales ou autres) fassent référence à certains aspects du premier événement, nous nous trouverions confrontés à une question gênante : est-ce que ces événements n'ont jamais été perçus consciemment, ou bien est-ce qu'ils ont été rapidement oubliés ?

Beaucoup d'expériences ont mesuré le « moment d'appréhension ». Dans un test de mémoire acoustique, vous entendez un enregistrement de nombreuses choses qui vous sont rapidement

présentées (par exemple quatre choses par seconde) et on vous demande de les identifier : vous ne pouvez tout simplement pas répondre tant que l'événement acoustique n'est pas achevé, et vous en identifiez certaines, mais pas les autres. Et pourtant subjectivement vous les avez toutes entendues aussi bien. La question qu'on peut se poser naturellement est la suivante : de quoi exactement étiez-vous conscients ? Il ne fait pas de doute que l'information sur la bande magnétique a été traitée par votre système auditif, mais est-ce que les marques identificatoires des entités non nommées ensuite vont directement vers votre conscience, ou bien est-ce qu'elles ont été enregistrées inconsciemment ? Elles semblent avoir été là dans la conscience, mais l'étaient-elles *réellement* ?

Dans un autre cas paradigmatique expérimental, on vous montre rapidement une diapositive sur laquelle sont imprimées de nombreuses lettres. (On fait cela avec un *tachistoscope*, un dispositif que l'on peut ajuster exactement de manière à présenter un stimulus d'une certaine luminosité pendant un certain nombre de millisecondes – quelquefois seulement 5 msec, quelquefois 500 msec, ou un intervalle plus long.) Vous pouvez plus tard rapporter avoir identifié seulement quelques-unes des lettres, mais vous avez certainement vu les autres. Vous insistez sur le fait qu'elles étaient vraiment là ; vous savez exactement combien il y en avait, et vous avez l'impression qu'elles étaient nettes et bien distinctes. Et pourtant vous ne pouvez pas les identifier. Est-ce que vous les avez rapidement oubliées ou bien ne les avez-vous jamais consciemment perçues à l'origine ?

Le phénomène bien étudié du *métacontraste* (Fehrer et Raab, 1962) introduit avec acuité l'idée principale du modèle des Versions Multiples. (Des phénomènes similaires sont exposés par Breitmeyer, 1984.) Si un stimulus est envoyé brièvement sur un écran (disons, pendant 30 msec – à peu près le même temps qu'une image télévisée) puis est immédiatement suivi par un second stimulus qui le « masque », les sujets *rapportent* qu'ils n'ont vu que le second stimulus. Le premier stimulus peut être un disque coloré et le second un anneau coloré qui s'adapte étroitement à l'espace du disque.

Si vous vous mettiez à la place du sujet, vous verriez vous-même ; vous seriez prêt à jurer qu'il n'y avait qu'un seul stimulus : l'anneau. Dans la littérature psychologique, la descrip-

tion standard de ce genre de phénomènes est stalinienne : le second stimulus, d'une certaine façon, *empêche l'expérience consciente* du premier. Autrement dit, il prend en quelque sorte le relais du premier stimulus dans son chemin vers la conscience. Les gens peuvent pourtant faire mieux que s'en remettre au hasard si on leur demande de deviner s'il y avait un ou deux stimuli. Cela montre seulement, encore une fois, comme dit la théorie stalinienne, que des stimuli peuvent avoir leurs effets sur nous sans que nous en soyons conscients. Le premier stimulus ne joue jamais sur la scène de la conscience, mais a des effets entièrement inconscients. Nous pouvons contrer cette explication du métacontraste avec l'autre hypothèse, l'hypothèse orwellienne : les sujets sont bien conscients du premier stimulus (ce qui explique la capacité qu'ils ont de les deviner correctement), mais leur mémoire de cette expérience consciente est *presque* entièrement oblitérée par le second stimulus (ce qui explique pourquoi ils nient l'avoir vu, en dépit de leurs soi-disant intuitions supposées meilleures que se fier au hasard). Le résultat conduit à un blocage, et c'est ennuyeux pour les deux parties en présence, parce que aucune d'elles ne peut identifier un résultat expérimental quelconque qui *permettrait* de résoudre le conflit.

Voici comment le modèle des Versions Multiples traite le méta-contraste. Quand beaucoup de choses arrivent en un temps très court, le cerveau peut forger des hypothèses simplificatrices. Le contour externe d'un disque tourne rapidement dans les contours

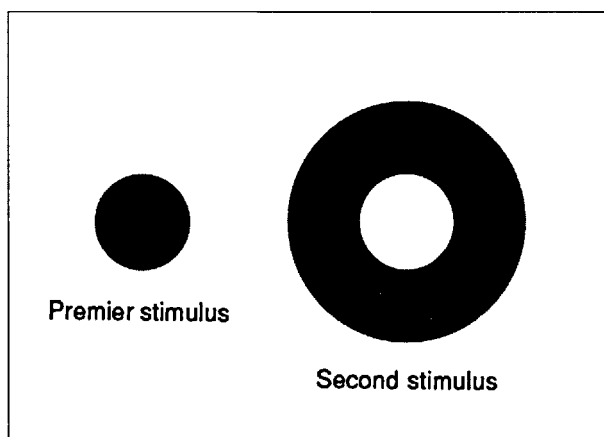


Figure 5.1

internes d'un anneau. Le cerveau, initialement informé seulement du fait que quelque chose s'est produit (quelque chose avec un contour particulier dans un lieu particulier), reçoit rapidement la confirmation du fait qu'il y avait effectivement un anneau, avec un contour interne et un contour externe. Sans disposer des données supplémentaires sur le fait qu'il y avait un disque, le cerveau arrive à la conclusion conservatrice qu'il n'y avait qu'un anneau. Devrions-nous insister sur le fait que le disque a été l'objet d'une expérience seulement parce que *si l'anneau n'était pas intervenu*, le disque aurait été rapporté ? Cela aurait été une erreur seulement consistant à supposer que nous puissions « geler » le film dans le Théâtre Cartésien et nous assurer que l'image du disque est réellement entrée dans le Théâtre avant que le souvenir de celui-ci n'ait été oblitéré par des événements ultérieurs. Le modèle des Versions Multiples s'accorde pour dire que l'information portant sur le disque s'est trouvée brièvement dans une position fonctionnelle qui lui permette de contribuer à un compte rendu ultérieur, mais que cet état s'est ensuite estompé ; il n'y a aucune raison de soutenir que cet état se trouvait dans le cercle enchanté de la conscience jusqu'à ce qu'il soit réécrit ou au contraire qu'il n'a jamais atteint cet état privilégié. Des versions composées à différents moments et dans différents lieux dans le cerveau peuvent avoir été sélectionnées pour constituer le contenu de la *conscience*.

Une démonstration encore plus frappante de cette capacité de révision est *le lapin cutané*. Les psychologues Frank Geldard et Carl Sherrick ont rapporté l'expérience initiale en 1972 (voir aussi Geldard, 1977 ; Geldard et Sherrick, 1983, 1985). Le bras du sujet reste au repos sur une table ; on place des frappeurs mécaniques à deux ou trois endroits le long du bras, jusqu'à trente centimètres les uns des autres. Les frappeurs envoient une série de coups, par exemple cinq dans le poignet, suivis de deux près du coude et de trois autres encore sur la partie supérieure du bras. Les coups sont donnés avec des intervalles entre chaque stimulus compris entre 50 et 200 msec. Par conséquent, une suite de coups peut durer moins d'une seconde, ou deux ou trois secondes au plus. Cela produit un effet étonnant : les coups semblent au sujet voyager en suites régulières sur des points équidistants réguliers jusqu'en haut du bras – comme si un petit animal tressautait le long du bras. Au départ, on se demande *comment le cerveau* a bien pu savoir qu'après cinq coups sur le poignet, il devait y avoir d'autres coups près du coude. Les sujets

font l'expérience du « départ » des coups du poignet à partir du second coup, et pourtant, dans des essais rapides où les coups ultérieurs dans le coude ne sont jamais donnés, ils éprouvent les cinq coups sur le poignet de la façon attendue. Le cerveau ne peut évidemment pas « savoir » quoi que ce soit sur un coup dans le cerveau avant qu'il ne se produise. Si vous êtes encore sous le charme de l'idée du Théâtre Cartésien, il se peut que vous vouliez imaginer que le cerveau retarde l'expérience consciente jusqu'à un moment suivant à celui où les coups ont été reçus à une certaine étape intermédiaire entre le bras et le siège de la conscience (quel qu'il soit), et que cette étape intermédiaire révise les données de manière à s'accorder avec une théorie du mouvement, et envoie la version éditée vers la conscience. Mais le cerveau retarderait-il toujours sa réponse à un coup s'il y en avait plus d'un ? Si ce n'est pas le cas, comment « sait »-il quand il lui faut retarder ?

Le modèle des Versions Multiples montre que cette question est mal posée. Le déplacement spatial (le long du bras) fait l'objet d'une discrimination temporelle par le cerveau. Le nombre de coups fait aussi l'objet d'une discrimination. Bien que dans la réalité physique, les coups soient regroupés à un certain endroit, une hypothèse simplificatrice consiste à dire qu'ils ont été regroupés régulièrement le long de l'extension spatio-temporelle de l'expérience. Le cerveau s'autorise cette interprétation parcimonieuse mais fausse, *après* que les coups ont été enregistrés, bien entendu, ce qui a pour effet de balayer les interprétations antérieures (partiales) des coups reçus, bien que des effets secondaires de ces interprétations puissent continuer à se produire. Par exemple, supposez que nous demandions aux sujets de presser un bouton chaque fois qu'ils ressentent *deux coups sur le poignet* : il ne serait pas surprenant qu'ils puissent commencer à presser le bouton *avant* que les coups sur l'avant-bras n'aient été discriminés et ne les aient induits à interpréter à tort le second coup comme se déplaçant sur le haut du bras.

Nous devons faire particulièrement attention de ne pas commettre l'erreur de supposer que le contenu que nous dériverions d'un sondage aussi précoce constitue le « premier chapitre » du contenu que nous trouverions dans le récit si nous devions sonder le même phénomène plus tard. Cela revient à confondre deux « espaces » différents, l'espace du mode de repré-

sensation et l'espace représenté. C'est une erreur si tentante et si fréquente qu'elle mérite qu'on lui consacre une section.

Comment le cerveau représente le temps

Le matérialisme cartésien, thèse que personne n'admet mais qui tend à influencer la pensée de tout le monde, suggère l'image implicite suivante. Nous savons que l'information se déplace dans le cerveau et est traitée par divers mécanismes dans diverses régions. Nos intuitions nous suggèrent que notre flux de conscience consiste en événements qui surviennent par séquences, et qu'à chaque instant, tout élément dans cette séquence peut être rangé au nombre de ceux qui ont déjà eu lieu « dans la conscience » ou de ceux qui n'ont pas encore eu lieu « là ». Si c'est le cas, alors il semble que les véhicules qui contiennent des contenus et se meuvent dans le cerveau doivent être comme des wagons sur une voie ; l'ordre dans lequel ils passent à un certain point sera celui dans lequel ils « arrivent » au théâtre de la conscience et (par conséquent) « deviennent conscients ». Pour déterminer où dans le cerveau apparaît la conscience, tracez toutes les trajectoires des véhicules d'information, et voyez par quel point des véhicules particuliers passent à l'instant où ils deviennent conscients.

Une réflexion sur la tâche fondamentale du cerveau nous montrera où cette image pêche. La tâche du cerveau consiste à guider le corps, qu'il contrôle à travers un monde de conditions changeantes et de surprises soudaines, en sorte qu'il doit récolter de l'information en provenance de ce monde et l'utiliser *rapidement* pour « produire du futur », pour extraire des anticipations afin d'être toujours un pas au-delà du désastre (Dennett, 1984a, 1991b). Le cerveau doit donc représenter des propriétés temporelles des événements du monde, et ce de manière efficace. Les processus responsables de l'exécution de cette tâche sont spatialement distribués dans un vaste cerveau sans nœud central, où la communication est relativement lente entre les régions ; les impulsions nerveuses électrochimiques voyagent des milliers de fois plus lentement que la lumière (ou que des signaux électroniques dans des fils). Par conséquent, le cerveau subit une pression temporelle significative. Il doit souvent s'arranger pour moduler sa sortie à la lumière de son entrée à l'intérieur d'une séquence temporelle

qui ne lui autorise pas de retards. Du côté des entrées, des tâches d'analyse perceptuelle, comme la perception du discours parlé, iraient bien au-delà des limites physiques de la machinerie du cerveau si elles n'utilisaient pas d'ingénieuses stratégies d'anticipation reposant sur des redondances de l'entrée. Le discours parlé normal a lieu à une vitesse de quatre ou cinq syllabes par seconde, mais les machines dont l'évolution nous a pourvus pour l'« analyser » (*parse*) sont si puissantes que les gens sont capables de comprendre du « discours compressé » dans lequel les mots sont envoyés à toute vitesse électroniquement sans produire un ton accéléré — à des vitesses qui vont jusqu'à trente syllabes par seconde. Du côté de la sortie, beaucoup d'actes doivent se produire si rapidement et être déclenchés avec une précision telle que le cerveau n'a pas le temps d'ajuster ses signaux de contrôle à la lumière du feed-back. Des actes tels que jouer au piano ou envoyer un caillou avec précision (Calvin, 1983, 1986) doivent être *balistiquement* enclenchés. (Les actions balistiques sont des missiles *non* guidés ; une fois qu'ils sont enclenchés, on ne peut pas ajuster leurs trajectoires.)

Comment, par conséquent, le cerveau s'y prend-il pour suivre à la trace l'information dont il a besoin ? Considérons le problème suivant : puisque la distance entre la hanche et le cerveau est bien plus grande que la distance entre le genou et le cerveau, ou que la distance entre le front et le cerveau, des stimuli produits simultanément à ces sites différents arriveront au Quartier Général en succession par saccades, si la vitesse de voyage est constante sur toutes les voies. Comment, pourriez-vous vous demander, le cerveau assure-t-il « une simultanéité centrale de représentation pour des stimuli distalement simultanés » ? En vous engageant dans un travail d'ingénierie spéculative à rebours, vous pourriez penser comme suit : peut-être tous les tracés des nerfs afférents sont-ils comme des mesures de bandes magnétiques réenroulées avec un ressort, et tous de même longueur : les nerfs qui vont vers les orteils sont complètement réenroulés, ceux qui vont vers le front sont essentiellement embobinés dans le cerveau. Des signaux intervenant dans les tracés du second type décrivent des cycles permanents dans leurs attaches internes, excitant le Quartier Général exactement au même instant que les signaux non cycliques qui viennent des orteils. Vous pouvez aussi imaginer que les tracés des nerfs diminuent de diamètre quand ils se tendent (comme des rouleaux en terre glaise ou des nouilles faites

à la maison) et que la vitesse de conduction varie avec le diamètre. (C'est ce qui se produit effectivement, mais malheureusement dans la mauvaise direction ! Les fibres plus épaisses ont une vitesse de conduction plus rapide.) Ce sont là des modèles propres à stimuler l'imagination (bien qu'ils soient stupides) et destinés à résoudre ce problème, mais l'erreur se situe en amont : elle consiste à supposer que le cerveau *doit* résoudre ce problème. Le cerveau ne devrait pas résoudre ce problème, pour une raison évidente d'ingénierie : il gaspille un temps précieux en concédant à toute la gamme de ses opérations un programme pour le cas « le pire ». Pourquoi des signaux vitalement importants venus du front (par exemple) traînent-ils dans l'antichambre parce qu'il pourrait un jour y avoir une occasion au moment où des signaux concurrents venus des orteils auraient besoin de converger d'une manière ou d'une autre avec eux ² ?

Les ordinateurs digitaux ont eux aussi besoin de retards de ce genre pour rendre possibles les cas défavorables et assurer la synchronie. Le mécanisme dans un circuit parallèle d'addition qui retient les additions complètes jusqu'à ce qu'une pulsation de modulation temporelle les relâche est un dispositif proche de celui des nerfs cycliques qu'on vient d'imaginer. Et ceux qui construisent des superordinateurs doivent être extraordinairement attentifs pour s'assurer que les circuits qui connectent diverses parties sont de la même longueur, ce qui requiert souvent des boucles supplémentaires de fils. Mais les ordinateurs digitaux peuvent se permettre ce genre d'inefficacité locale parce qu'ils ont de la vitesse à dépenser. (En fait, avec la concurrence qui pousse à construire des ordinateurs digitaux de plus en plus rapides, ces petits défauts temporels sont à présent repensés ; la principale raison pour laquelle bon nombre d'entre eux demeurent est que les ingénieurs ne savent pas comment construire des systèmes informatiques totalement asynchrones, non régulés par une pulsation d'horloge centrale.)

Imposer une synchronie centrale sur des opérations requiert des retards. En nous faisant ingénieurs à rebours, nous pouvons spéculer que si le cerveau dispose de manières efficaces de représenter l'information sur le temps dont il a besoin de façon à éviter ces retards, l'évolution les aura « trouvées ». En fait, ces modes de représentation existent bien et nous pouvons le voir à l'aide d'un incident historique qui manifeste ce phénomène à grande échelle, à la fois dans l'espace et dans le temps.

Considérons les difficultés de communication que rencontrait, en son extension, l'Empire britannique avant l'invention de la radio et du télégraphe. Contrôler un empire étendu jusqu'aux quatre coins du monde à partir de quartiers généraux situés à Londres n'était pas toujours possible. L'incident le plus célèbre est sûrement la bataille de La Nouvelle-Orléans, le 8 janvier 1815, quinze jours après que le traité qui mettait fin à la guerre de 1812 a été signé en Belgique. Plus d'un millier de soldats britanniques furent tués dans cette bataille inutile. Nous pouvons prendre exemple sur cette débâcle pour voir comment le système fonctionnait. Supposez qu'au jour 1, le traité soit signé en Belgique ; la nouvelle est envoyée par terre et par mer vers l'Amérique, l'Inde, l'Afrique et ainsi de suite. Au quinzième jour, la bataille a lieu à La Nouvelle-Orléans et la nouvelle de la défaite est envoyée par terre et par mer vers l'Angleterre, l'Inde, etc. Au vingtième jour, donc trop tard, la nouvelle du traité (et l'ordre de se rendre) arrive à La Nouvelle-Orléans. Au trente-cinquième jour, supposons, la nouvelle de la défaite arrive à Calcutta, mais la nouvelle du traité n'arrive pas là-bas avant le quarantième jour (elle voyage par terre, donc lentement). Aux yeux du commandant en chef britannique en Inde, la bataille « semblerait » avoir eu lieu avant la signature du traité – mais l'habitude de dater les lettres lui permet d'effectuer les corrections nécessaires ³.

Les agents éloignés résolvaient la plupart de leurs problèmes de communication temporelle en incluant des modes de représentations du temps pertinent de l'information dans les *contenus* de leurs signaux, en sorte que le temps d'arrivée des signaux eux-mêmes soit *strictement non pertinent* pour l'information qu'ils véhiculaient. Une date écrite en haut d'une lettre (ou un tampon postal sur l'enveloppe) donne au récipiendaire l'information nécessaire sur le moment où elle a été envoyée : l'information survit à tout retard dans la transmission ⁴. Cette distinction entre le temps marqué (sur le tampon postal) et le temps du mode de représentation (quand la lettre arrive) est un exemple de la distinction familière entre le contenu et le véhicule. Alors que les dispositifs de communication du cerveau n'ont pas à leur disposition cette solution particulière (parce qu'ils ne « savent pas la date » quand ils envoient leurs messages), le principe général de la distinction contenu/véhicule est pertinent pour les modèles de traitement de l'information dans le cerveau, bien qu'on ne s'en soit pas souvent rendu compte ⁵.

En général, on doit distinguer les propriétés des *modes* de représentation et celles des *choses* représentées. Quelqu'un peut crier à tue-tête : « Tout doux, marchez sur la pointe des pieds ! » Il existe des images gigantesques représentant des objets microscopiques et une peinture à l'huile représentant un artiste en train de faire une esquisse au fusain n'a rien d'impossible. La première phrase de la description écrite d'un homme debout n'a pas besoin de décrire sa tête, et la dernière phrase n'a pas besoin de décrire ses pieds. Ce principe s'applique également, mais de façon moins évidente, au temps. Considérons *l'expression parlée* « un flash brillant et rapide de lumière rouge ». Le début de *cette* expression est « un flash » et sa fin est « de lumière rouge ». Ces portions de cet événement discursif ne sont pas elles-mêmes des représentations de débuts ou de fins de flashes rouges rapides (Efron, 1967, p. 714, fait la même remarque). Aucun événement du système nerveux ne peut avoir de durée nulle (pas plus qu'il ne peut avoir une extension spatiale nulle) ; par conséquent, son début et sa fin sont séparés par un certain intervalle temporel. Si l'événement lui-même *représente* un événement dans l'expérience, alors l'événement qu'il représente doit lui-même avoir une durée non nulle, un début et une fin. Mais il n'y a pas de raison de supposer qu'au départ le mode de représentation représente le début du représenté ⁶. Bien que différents attributs soient bien entendu extraits par différents dispositifs neuronaux à différentes vitesses (par exemple la localisation forme-couleur), et bien que si on nous demandait de répondre à la présence de chacune d'elles isolément, nous le ferions avec différents temps de latence, nous percevons les événements, et non un filet d'éléments perceptifs ou d'attributs analysés au fur et à mesure ⁷.

Un roman ou un récit historique peut n'avoir pas été composé dans l'ordre qu'il finit par représenter – il arrive que les auteurs commencent par la fin et remontent à rebours dans le temps. De plus, un tel récit peut contenir des flash-back, dans lesquels les événements sont *représentés comme* étant arrivés dans un certain ordre par le moyen de *modes de représentation* qui surviennent dans un ordre différent. De même, la représentation par le cerveau de *A avant B* n'a pas besoin d'être accomplie par en premier lieu :

une représentation de A

suivie de :

une représentation de B.

L'expression « B après A » est un exemple de véhicule (parlé) qui représente A avant B ; le cerveau peut se permettre la même liberté de placement temporel. Ce qui compte pour le cerveau, ce n'est pas nécessairement le moment où des événements individuels représentants surviennent dans différentes parties du cerveau (du moment qu'ils arrivent à temps pour contrôler les choses qui ont besoin de l'être !), mais leur *contenu temporel*. Autrement dit, ce qui compte, c'est que la possibilité pour le cerveau de se mettre à contrôler des événements « sous l'hypothèse que A a eu lieu avant B » que l'information que A est arrivé soit entrée dans les systèmes pertinents et ait été reconnue comme telle avant ou après l'information que B a eu lieu. (Rappelez-vous le commandant en chef à Calcutta : tout d'abord, il fut informé de la bataille, puis il fut informé du traité, mais parce qu'il peut extraire de cela l'information que le traité a eu lieu avant, il peut agir en conséquence. Il doit *juger* que le traité a eu lieu avant la bataille ; il ne doit pas en plus construire une sorte de tableau de « reconstruction historique », dans lequel il reçoit les lettres « dans le bon ordre ».)

Certains auteurs ont soutenu, cependant, que le temps est ce que l'esprit ou le cerveau *doit* représenter « avec lui-même ». Le philosophe Hugh Mellor, dans son livre *Real Time* (1981, p. 8), l'affirme clairement et vigoureusement :

« Supposez par exemple que je voie un événement *e* en précéder un autre, *e**. Je dois d'abord voir *e* puis *e**, ma vision de *e* étant en quelque sorte mémorisée dans ma vision de *e**. C'est-à-dire que ma vision de *e* affecte ma vision de *e** : c'est ce qui me fait, à tort ou à raison, voir *e* précéder *e** plutôt que l'inverse. Mais voir *e* précéder *e** signifie voir *e* en premier. Par conséquent, l'ordre causal de mes perceptions de ces événements, en fixant l'ordre temporel que je perçois en eux, fixe l'ordre temporel des perceptions elles-mêmes [...] Le fait frappant [...] doit être relevé, à savoir que les perceptions de l'ordre temporel ont besoin de perceptions temporellement ordonnées. *Aucune autre propriété ou relation n'a ainsi à être incorporée dans les perceptions qu'on en a* (mes italiques) : les perceptions de forme et de couleur, par exemple, n'ont pas besoin en elles-mêmes d'être semblablement douées de forme ou de couleur. »

C'est faux, mais il y a là quelque chose de vrai. Étant donné que la fonction principale de représentation dans le cerveau est le contrôle du comportement en temps réel, l'ordre temporel des

représentants est *jusqu'à un certain point* essentiel à leur tâche, et ce de deux manières.

En premier lieu, au début d'un processus perceptuel, l'ordre temporel peut être *ce qui détermine le contenu*. Considérons la façon dont nous distinguons un point en train de se mouvoir de gauche à droite sur un écran de cinéma. La *seule* différence entre les deux peut être l'ordre temporel dans lequel deux images (ou plus) sont projetées. Si A est d'abord projeté, puis B, on voit le point se mouvoir dans une direction. La seule différence dans les stimuli que le cerveau puisse utiliser pour discriminer en cette direction est l'ordre dans lequel ils arrivent. Cette discrimination est donc, logiquement, fondée sur la capacité du cerveau à faire une discrimination d'ordre temporel d'une certaine acuité. Puisque les images filmiques sont habituellement envoyées à une vitesse de vingt-quatre par seconde, nous savons que le système visuel peut établir un ordre entre les stimuli qui surviennent pendant environ 50 msec. Cela veut dire que les propriétés temporelles effectives des signaux — leurs points de départ, leur vélocité dans le système, et par conséquent leur temps d'arrivée — doivent être contrôlées de façon précise jusqu'à ce que soit faite une telle discrimination. Autrement, l'information sur laquelle la discrimination doit être fondée sera perdue ou brouillée.

À plus grande échelle, ce phénomène survient quelquefois au début d'une course de voiliers ; vous *voyez* un bateau franchir la ligne de départ et vous *entendez* ensuite le coup de feu signalant le départ. Mais le bateau n'était-il pas sur la ligne trop tôt ? C'est logiquement impossible à dire à moins que vous ne calculiez les différents temps de transmission du son et de la lumière vers le lieu où vous avez fait la discrimination. À partir du moment où un jugement a été fait (soit *tout va bien*, soit *le bateau numéro 7 était trop tôt sur la ligne*), ce contenu peut être transmis aux participants de manière nonchalante, sans qu'il soit besoin de s'occuper de sa vitesse ou de la distance qu'il doit parcourir pour accomplir sa tâche.

Par conséquent, l'ordre temporel de certains représentants compte *jusqu'au moment où* on effectue une discrimination telle que *de gauche à droite* (ou *sur la ligne trop tôt*), mais à partir du moment où elle a été faite, localement, par un circuit quelconque dans le cortex (ou par certains observateurs sur le bateau du comité), le contenu du jugement peut être envoyé, avec un

relâchement temporel, n'importe où dans le cerveau où cette information pourrait être utilisée. C'est seulement de cette manière que l'on peut expliquer le fait suivant, qui, autrement, serait problématique : les gens sont parfois incapables de réussir par hasard certains jugements sur un ordre temporel quelconque, alors qu'ils réussissent sans problème à produire d'autres jugements (comme des jugements sur la direction du mouvement) qui requièrent logiquement bien plus d'acuité temporelle. Ils utilisent des discriminateurs spécialisés pour produire des jugements fins.

La seconde contrainte pesant sur l'ordre temporel a déjà été signalée : l'ordre des représentants ne compte pas tant qu'ils arrivent suffisamment à temps pour contribuer au contrôle du comportement approprié. La fonction d'un représentant peut dépendre du respect d'un *délai*, propriété temporelle du véhicule qui produit la représentation. Cela apparaît avec une netteté particulière dans des environnements soumis à la pression du temps tels que l'Initiative de Défense Stratégique imaginée récemment. Le problème n'est pas dans ce cas de savoir comment on peut faire en sorte qu'un système informatique représente, avec précision, des lanciers de missiles, mais de savoir comment représenter avec précision un lancer de missile pendant le bref intervalle pendant lequel vous pouvez encore le diriger. Un message disant qu'un missile a été lancé à 6:04:23.678 du matin EST peut représenter précisément et pour toujours le temps du lancer, mais son utilité peut s'évanouir complètement à 6:05 du matin EST. Pour toute tâche de contrôle, par conséquent, il existe une *séquence de contrôle temporel* pendant laquelle les paramètres temporels des représentants peuvent en principe être changés *ad libitum*.

Les délais qui limitent de telles séquences ne sont pas fixés ; ils dépendent plutôt de la tâche accomplie. Si, au lieu d'intercepter des missiles, vous écrivez vos mémoires ou répondez à des questions pendant les auditions du Watergate (Neisser, 1981), l'information que vous avez besoin de retrouver au sujet de la suite d'événements de votre vie pour contrôler vos actions peut être retrouvée dans à peu près n'importe quel ordre ; vous pouvez prendre le temps de faire des inférences. Ou, pour donner un exemple intermédiaire plus proche des phénomènes que nous sommes en train d'analyser, supposez que vous dériviez dans un bateau et que vous vous demandiez si vous dérivez droit devant

ou à distance d'un récif dangereux que vous voyez au loin. Supposez que *maintenant* vous connaissiez la distance effective qui vous sépare du récif (par exemple en mesurant l'angle qu'il fait par rapport à votre champ visuel) ; afin de répondre à votre question, vous pouvez attendre un petit peu, puis mesurer l'angle à nouveau ; si une demi-heure auparavant vous avez pris une photo polaroïd du récif, vous pouvez aussi mesurer l'angle sur cette vieille photo, faire certains calculs, et *rétrospectivement* vous figurer à quelle distance vous étiez alors. Afin de juger de la direction vers laquelle vous dérivez, il vous faut calculer deux distances : une distance à midi, et une distance à 12:30, par exemple, mais le fait de calculer une distance ou une autre en premier n'a pas d'importance. Cependant, vous avez intérêt à pouvoir faire le calcul assez rapidement de façon à sortir les rames avant qu'il ne soit trop tard.

Par conséquent, la représentation du temps par le cerveau est ancrée au temps lui-même de deux manières : l'ordre même du représentant peut fournir les données ou déterminer le contenu, et la raison même pour laquelle on représente le temps des choses peut être perdue si le représentant ne se produit pas dans le temps pour faire la différence qu'il est supposé faire. J'espère que Mellor prend en compte ces deux facteurs et qu'il les avait à l'esprit lorsqu'il a avancé la thèse que j'ai citée. Mais il commet l'erreur naturelle qui consiste à penser que ces facteurs combinés conditionnent entièrement la représentation du temps, en sorte que cet ordre de représentation représente *toujours* l'ordre du contenu. Selon son analyse, il n'y a pas place pour un « étalement » temporel, alors qu'il *devait* y avoir un étalement temporel – sur une petite échelle – *parce qu'il doit* y avoir un étalement spatial (sur une petite échelle) du point de vue de l'observateur.

Les causes doivent précéder les effets. Ce principe fondamental assure que les séquences de contrôle temporel sont limitées aux deux bouts de la chaîne : par le premier moment où l'information pourrait arriver dans le système et par le dernier moment où l'information pourrait contribuer causalement au contrôle d'un comportement particulier. Nous n'avons toujours pas vu *comment* le cerveau pourrait utiliser le temps disponible dans une séquence temporelle pour trier l'information qu'il reçoit et la transformer en un « récit » cohérent qui peut être utilisé pour gouverner les réponses du corps.

Comment, par conséquent, des propriétés temporelles peuvent-

elles être inférées par des processus dans le cerveau ? En théorie, des systèmes de « cachets postaux » ou de « tampons postaux » ne sont pas impossibles, mais il existe une méthode moins coûteuse, moins sûre mais biologiquement plus plausible : elle suppose ce que nous pourrions appeler un *dispositif sensible au contenu* et comparable au studio de cinéma où la bande-son est synchronisée avec le film. Les divers segments de la bande-son peuvent eux-mêmes avoir perdu leurs marqueurs temporels, en sorte qu'on ne peut pas simplement, mécaniquement, les coordonner avec les images. Mais en les passant dans un sens ou dans un autre relativement au film et en cherchant les convergences, on parvient en général à un ajustement « idoine ». Le petit clap sur la bande au début de chaque prise – « scène trois, prise sept, caméra en marche, *clap !* » – fournit un double repère, un clap auditif et un clap visuel, qui doivent glisser synchroniquement, en poussant le reste de la bande et les images en position au même moment. Mais les points de repère mutuels sont si nombreux que ce repère conventionnel au début de chaque prise n'est qu'une redondance commode. Obtenir un bon enregistrement dépend du *contenu* du film et de la bande, mais pas d'une analyse sophistiquée du contenu. Un éditeur de films qui ne connaîtrait pas le japonais trouverait que synchroniser une bande-son japonaise sur un film japonais est difficile et barbant, mais possible. De plus, l'ordre temporel des étapes du processus consistant à enregistrer les divers morceaux est indépendant du contenu du produit ; l'éditeur peut organiser la scène trois avant d'organiser la scène deux, et en principe, il pourrait même faire tout le boulot en passant les segments « à l'envers ».

Des processus tout à fait « stupides » peuvent réaliser des manipulations et des découpages dans le cerveau. Par exemple, le calcul de la profondeur dans des stéréogrammes des points au hasard (voir figure 4.7, p. 147) est un problème spatial, dont nous pouvons imaginer divers analogues temporels. En principe, par conséquent, le cerveau peut résoudre certains de ses problèmes d'inférence temporelle par l'intermédiaire d'un tel processus, en tirant des données non pas des yeux gauche et droit, mais de n'importe quelles sources d'information impliquées dans un processus requérant des jugements temporels.

On peut en tirer deux conséquences importantes. En premier lieu, on peut former des inférences temporelles (ces discriminations temporelles sont possibles) en comparant le *contenu* (au

niveau inférieur) de diverses gammes de données, et ce processus en temps réel n'a pas besoin de se produire dans l'ordre temporel que finit par représenter son produit. En second lieu, à partir du moment où une inférence temporelle a été accomplie – ce qui peut avoir lieu *avant* que des propriétés temporelles aient été extraites par d'autres processus –, il n'est pas nécessaire qu'elle soit recommencée ! Il n'est pas nécessaire qu'il existe une représentation *ultérieure* dans laquelle les propriétés de niveau supérieur seraient « présentées » dans une séquence en temps réel pour le bénéfice d'un second dispositif chargé de juger la séquence en question. En d'autres termes, après avoir formé ces inférences à partir de ces juxtapositions d'information temporelle, le cerveau peut se mettre à représenter les résultats sous n'importe quel format qui convienne à ses besoins et à ses ressources – et pas nécessairement un format dans lequel « le temps est utilisé pour représenter le temps ».

Le « renvoi en arrière dans le temps » de Libet

Voilà comment le cerveau peut accomplir son travail éditorial sur l'information temporelle de telle façon qu'il ignore l'ordre temporel effectif (le « temps d'arrivée ») de certaines de ses représentations. Toutefois, il doit effectuer ces opérations sous une forte pression temporelle. En remontant à rebours à partir du délai requis, tout contenu reporté ou exprimé sous une autre forme dans un comportement ultérieur doit avoir été présent (dans le cerveau, pas nécessairement « dans la conscience ») au moment requis pour qu'il contribue causalement à ce comportement. Par exemple, si un sujet, dans une expérience, dit « chien » en réponse à un stimulus visuel, nous pouvons travailler à rebours à partir du comportement ; celui-ci était de toute évidence contrôlé par un processus dont le contenu était *chien* (à moins que le sujet n'ait dit « chien » en réponse à tous les stimuli, ou ne passe sa journée à dire « chien chien chien chien... » et ainsi de suite). Comme il faut 100 msec environ pour *commencer* à exécuter une intention de discours de ce type (et en gros 200 autres msec pour l'accomplir), il est certain que le contenu *chien* était présent dans (approximativement) les aires linguistiques du cerveau environ 100 msec avant que l'énoncé n'ait commencé. Si nous partons de l'autre bout de la chaîne, à

nouveau, nous pouvons déterminer le moment initial où le contenu *chien* a été calculé ou extrait par le système visuel à partir de l'entrée rétinienne, et même, peut-être, suivre sa création et sa trajectoire ultérieure à travers le système visuel et en direction des aires du langage.

Ce qui serait véritablement anormal (et source de plaintes et de grincements de dents), ce serait que le temps qui s'est écoulé entre le stimulus *chien* et l'énonciation « chien » soit plus petit que le temps physiquement requis pour que ce contenu soit établi et se soit mû à travers le système. Mais on n'a jamais observé ce genre d'anomalies.

Cependant, on a découvert certaines juxtapositions étonnantes entre les deux séquences dessinées sur la figure 4.12 de la page 175. Quand nous essayons de placer la suite d'événements du flux objectif de traitement dans le cerveau en accord avec la suite subjective du sujet *en tant que déterminée par ce que le sujet dit ensuite* nous découvrons quelquefois des recoupements étonnamment vastes. Telle est, du moins, la conclusion que nous pourrions vouloir tirer de l'une des expériences les plus discutées et les plus critiquées des neurosciences : l'expérience neurochirurgicale de Benjamin Libet démontrant ce qu'il appelle « un renvoi à rebours dans le temps ».

Il arrive que pendant une opération sur le cerveau il soit important que le sujet soit éveillé, alerte et soumis seulement à une anesthésie locale (comme quand on nous donne de la novocaïne chez le dentiste). Cela permet au neurochirurgien d'obtenir des comptes rendus immédiats de la part du patient quant à ce qu'il éprouve au moment où le cerveau est opéré (voir la note 5, p. 88). C'est Penfield (1958) qui a été le pionnier de ce genre de pratiques, et pendant plus de trente ans les neurochirurgiens ont récolté des données sur les résultats de stimulations électriques directes sur diverses parties du cortex. On sait depuis longtemps que des stimulations de localisations sur le cortex *somatosensoriel* (une bande localisée heureusement sur le sommet du cerveau) produisent chez le patient l'expérience de sensations sur des parties correspondantes du corps. Par exemple, la stimulation d'un point sur le cortex somatosensoriel gauche peut produire la sensation d'un bref pincement sur la main droite du sujet (en raison de l'inversion bien connue dans le système nerveux qui fait que la partie gauche du cerveau est responsable de la partie droite du corps, et vice versa). Libet a comparé le temps

mis par des pincements induits sur le cortex à celui de sensations produites de la façon usuelle, en appliquant une brève impulsion électrique à la main elle-même (Libet, 1965, 1981, 1982, 1985b ; Libet et al., 1979 ; voir aussi Popper and Eccles, 1977 ; Dennett, 1979b, Churchland, 1981a ; Honderich, 1984).

Que faut-il attendre ? Supposez que deux personnes se rendent à leur travail chaque jour exactement au même moment, mais que l'un habite dans une banlieue éloignée, alors que l'autre habite à quelques pâtés de maisons de son bureau. Ils conduisent à la même vitesse, en sorte qu'étant donné la distance supplémentaire que doit parcourir le banlieusard, nous pouvons compter qu'il arrive au bureau plus tard. Ce n'est pas, cependant, ce que Libet a découvert quand il a demandé à ses patients ce qui venait en premier, le pincement à la main qui a pris naissance dans le cortex ou le pincement à la main envoyé à partir de la main. À partir des données recueillies, il a soutenu qu'alors que dans chaque cas cela prenait un temps considérable (environ 500 msec) du départ de la stimulation jusqu'à l'« adéquation neuronale » (le point où selon lui les processus corticaux culminent pour produire une expérience consciente d'un pincement), quand la main elle-même était stimulée, l'expérience était « automatiquement » « renvoyée à rebours dans le temps » et était ressentie comme arrivant *avant* le pincement produit par la stimulation cérébrale elle-même.

Chose plus frappante encore, Libet a rapporté des cas dans

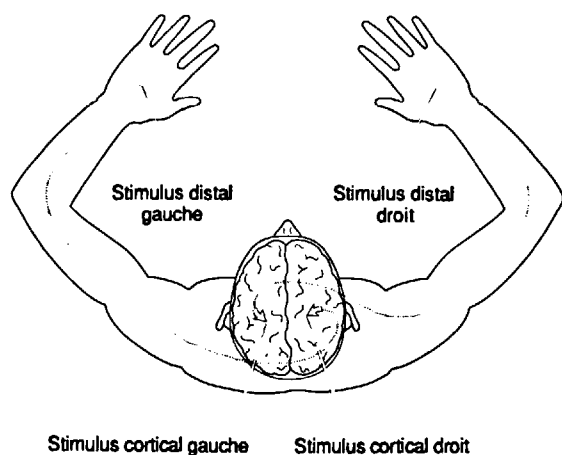


Figure 5.2

lesquels le *cortex* gauche d'un patient était stimulé *avant* que sa *main* gauche ne soit stimulée, ce qui, tendrait-on à penser, donnerait sûrement naissance à deux pincements ressentis : d'abord la main droite (corticalement induite) et puis ensuite la main gauche. En fait, cependant, le compte rendu subjectif était inversé : « D'abord gauche, ensuite droite. »

Libet a interprété ses résultats comme un défi sérieux à l'encontre du matérialisme : « Une dissociation entre les ordres temporels des événements correspondants " mentaux et physiques " semblerait poser des difficultés sérieuses, bien que surmontables pour [...] la théorie de l'identité psychoneuronale » (Libet et al., 1979, p. 22). Selon Sir John Eccles, lauréat du prix Nobel de médecine pour sa recherche en neurophysiologie, ce défi ne peut être relevé :

« La procédure d'antidatage ne semble pas être explicable par un processus neurophysiologique quelconque. On peut présumer que c'est une stratégie qui a été apprise par l'esprit conscient de soi... L'expérience sensorielle d'antidatage est attribuable à l'aptitude qu'a l'esprit conscient à faire de petits ajustements temporels, i.e à jouer des tours au temps » (Popper et Eccles, 1977, p. 364).

Plus récemment, le mathématicien et physicien Roger Penrose (1989) a suggéré qu'une explication matérialiste des phénomènes de Libet exigerait une révolution en physique fondamentale. Bien que l'expérience de Libet ait été tenue pour une large part dans les cercles non scientifiques comme une démonstration de la vérité du dualisme, peu de gens dans la communauté des sciences cognitives partagent cette opinion. En premier lieu, les procédures expérimentales de Libet, et son analyse des résultats, ont été sévèrement critiquées. Son expérience n'a jamais été reproduite, ce qui, dans de nombreux cercles, est une raison suffisante pour ne pas considérer ses « résultats ». La thèse sceptique, par conséquent, est que ces phénomènes n'existent tout simplement pas. *Mais que devrait-on dire s'ils existaient effectivement ?* C'est précisément le genre de questions qu'un philosophe poserait. Mais en l'occurrence, il est d'autres raisons de la poser que les motivations philosophiques usuelles. Personne ne doute de l'existence de phénomènes plus simples que celui de ϕ pour les couleurs et du lapin cutané, et leur interprétation pose les mêmes problèmes. Ce serait de la myopie théorique que de se contenter de les rejeter pour des *raisons méthodologiques*, en ne contestant pas

les hypothèses d'arrière-plan suggérant que si les expériences de Libet *étaient* un jour convenablement reproduites, ce serait un jour sombre pour le matérialisme.

Tout d'abord, l'expérience de Libet ne prouverait pas du tout l'existence d'une anomalie quelconque si nous négligions la possibilité d'enregistrer les comptes rendus verbaux des sujets sur leurs expériences afin de les utiliser ensuite pour engendrer d'abord un texte et ensuite un monde hétérophénoménologique. Les bruits qu'ils font avec leurs dispositifs vocaux pendant ou après les expériences n'ont rien de paradoxal si on les traite simplement comme des phénomènes acoustiques. En aucun cas, les sons ne paraissent sortir des têtes avant que les lèvres ne bougent, pas plus que les mains ne bougent avant les événements cérébraux qui sont supposés les causer. Des événements n'arrivent pas non plus dans le cortex avant les stimuli qui sont considérés comme les produisant. Considérés strictement comme le comportement interne et externe d'un système de contrôle du corps biologiquement réalisé, les événements qui sont observés et qui font l'objet d'un enregistrement temporel dans les diverses expériences ne révèlent aucune violation apparente de la causalité mécanique quotidienne – du type de celle dont la physique galiléenne/newtonienne nous fournit le modèle standard approximatif.

Vous pourriez donc « faire disparaître les problèmes » en vous transformant en behavioriste bon teint et en refusant tout simplement de prendre au sérieux les comptes rendus introspectifs. Mais nous ne sommes pas des behavioristes bon teint ; nous voulons accepter le défi et donner un sens à ce que Libet appelle « un aspect phénoménologique premier de l'existence humaine en relation avec la fonction cérébrale » (1985, p. 534). Libet s'approche très près de l'hétérophénoménologie. Il dit : « Il est important de réaliser que ces renvois subjectifs et ces corrections ont apparemment lieu au niveau de la " sphère mentale " ; ils ne sont pas apparents, comme tels, dans les activités aux niveaux neuronaux » (1982, p. 241). Mais comme aucune référence à la phénoménologie n'est *neutre*, il doit rapporter l'anomalie à la « sphère mentale ». C'est là un petit pas qu'il est obligé de faire (car il *doit* dire cela s'il rejette le behaviorisme), mais c'est le premier pas sur la planche glissante qui mène au dualisme.

« Les comptes rendus que font les sujets de leurs diverses expériences... n'étaient pas des constructions théoriques, mais des obser-

vations empiriques... La méthode introspective peut avoir ses limites, mais on peut l'utiliser de façon appropriée dans le cadre de la science naturelle, et elle est absolument essentielle si l'on essaie d'obtenir certaines données expérimentales sur le problème esprit-corps » (1987, p. 785).

Les comptes rendus des sujets, même quand on les transforme en textes, sont, comme le dit Libet, des observations empiriques, mais *ce qu'ils rapportent*, les événements dans leurs mondes hétérophénoménologiques, sont en fait des constructions théoriques. Ils peuvent, comme le soutient Libet, être utilisés correctement à l'intérieur du cadre des sciences de la nature, mais *seulement* si vous les comprenez d'emblée comme des fictions théoriques.

Libet soutient que ses expériences de stimulation directe sur le cortex démontrent « deux facteurs temporels remarquables » :

1. *Il existe un délai important avant que les activités cérébrales*, produites par un stimulus sensoriel, *parviennent à l'adéquation neuronale* pour produire un résultat quelconque dans l'expérience sensorielle consciente.

2. Une fois que l'adéquation neuronale a été atteinte, l'*ordre temporel* de l'expérience *est (automatiquement)* renvoyé à rebours dans le temps, utilisant un « signal temporel » sous la forme d'une réponse initiale du cortex cérébral en direction du stimulus sensoriel (1981, p. 182).

Le « signal temporel » est la *première* production d'activité à apparaître dans le cortex (le potentiel évoqué primaire), qui survient seulement pendant 10 ou 20 msec après la stimulation de l'organe sensoriel périphérique. Libet suggère que le renvoi à rebours est toujours « en direction » du signal temporel.

Le modèle de Libet est stalinien : après le potentiel primaire évoqué, divers processus d'édition surviennent dans le cortex avant le moment de l'« adéquation neuronale », et c'est à ce moment-là que le film achevé se trouve projeté. Comment est-il projeté ? Ici l'analyse de Libet hésite entre une conception extrême et une conception plus modérée (cf. Honderich, 1984) :

1. *Projection à rebours* : il est quelquefois envoyé à rebours dans le temps dans un Théâtre Cartésien quelconque où il est projeté en synchronie avec les potentiels évoqués primaires (les potentiels évoqués primaires, en tant que « signaux tempo-

rels », ont un usage comparable à celui du clap dans le tournage des films, montrant au projecteur à quelle distance à rebours dans le temps il doit envoyer l'expérience).

2. *Le renvoi à rebours* : il est projeté dans le temps ordinaire, mais il porte quelque chose comme un cachet postal, rappelant au spectateur que ces événements doivent être compris comme s'étant produits un peu avant. (Dans ce cas, les potentiels évoqués primaires servent simplement de dates, qui pourraient être *représentées* sur l'écran cartésien par le titre : « la veille de la bataille de Waterloo » ou « New York, été 1942 ».)

Le terme utilisé par Libet lui-même est celui de *renvoi*, et il le justifie en nous rappelant le phénomène « connu et accepté depuis longtemps » du renvoi spatial, ce qui suggère une interprétation modérée.

« Le concept de renvoi subjectif à rebours dans le temps est un étrange concept et peut-être n'est-il pas clair quand on l'expose initialement. Mais il a un prédécesseur majeur : le concept connu et accepté depuis longtemps de renvoi subjectif dans la dimension spatiale. Par exemple, l'image subjective dont on fait l'expérience en réponse à un stimulus visuel a une configuration et une localisation spatiales subjectives qui sont très différentes de la configuration et de la localisation spatiales des activités neuronales qui donnent naissance à l'image ("à référence subjective") » (1981, p. 183 ; voir aussi Libet et al., 1979, p. 221 ; Libet, 1985b).

Cependant, il poursuit en concluant que le renvoi temporel pose des problèmes pour le matérialisme (« la théorie de l'identité psycho-neuronale », Libet et al., 1979, p. 222), ce qui montre ou bien qu'il pense que le renvoi spatial pose aussi ces problèmes, ou bien qu'il n'a pas compris sa propre défense de cette idée. Si le renvoi spatial – le fait que ce que nous voyons semble être en dehors, et non pas au-dedans, de nos cerveaux – pose un problème pour le matérialisme, pourquoi Libet suggère-t-il que son propre travail contient un argument nouveau et important en faveur du dualisme ? Il ne fait pas de doute que le fait du renvoi spatial à rebours est bien plus attesté que les renvois temporels qu'il a bien de la peine à démontrer. Il semble, cependant, que Libet ait une conception radicale (ou à tout le moins confuse) du renvoi spatial comme « projection » :

« Il existe des données expérimentales en faveur de la thèse selon laquelle la "sphère" subjective ou mentale pourrait en fait "rem-

plir " des lacunes temporelles ou spatiales. Sinon, comment pourrait-on envisager le contraste énorme dont l'existence est établie entre une image visuelle subjective et la configuration des activités neuronales qui donne naissance à l'expérience de l'image ? » (1981, p. 196).

Cela semble vouloir dire que le projecteur que Smythies ne parvenait pas à trouver dans le cerveau se cache dans la « sphère » mentale⁸.

Comment Libet prétend-il établir ses deux remarquables facteurs temporels ? « L'adéquation neuronale » qui, selon Libet, requiert jusqu'à 500 msec d'activité corticale est déterminée en regardant à quel moment, à partir d'une stimulation initiale, une stimulation corticale directe peut *interférer avec* la conscience rapportée ensuite. Au-delà de cet intervalle critique, un stimulus cortical direct serait reporté par le sujet comme étant une expérience *ultérieure*. (Arrivée trop tard pour être incorporée par la salle de rédaction pour l'« épreuve finale » de la première expérience du stimulus, elle apparaîtrait dans l'édition suivante.) Les données de Libet suggèrent une séquence d'édition terriblement variable : « Le stimulus cortical conditionnant pourrait être enclenché plus de 500 msec après l'influx sur la peau et modifier encore la sensation de la peau, bien que dans la plupart des cas des effets rétroactifs ne soient pas observés avec des intervalles S-C de plus de 200 msec » (1981, p. 185). Libet fait attention de définir l'adéquation neuronale en termes d'effets produits sur des comptes rendus verbaux réfléchis subséquents : « On demandait au sujet de rapporter ce qu'il sentait, en un espace de quelques secondes après la production de chaque paire de stimuli » (1979, p. 195). Il insiste sur le fait que « l'ordre temporel d'une expérience subjective doit être distingué de celui d'une réponse comportementale (comme dans le temps de réaction), qui peut être faite avant que la connaissance conscience se développe » (Libet et al., 1979, p. 193).

Cette restriction lui permet de défendre une interprétation opposée aux données établies par Patricia Churchland. Churchland est la première « neurophilosophe » (voir son livre de 1986, *Neurophilosophy : Toward a Unified Science of the Mind-Brain*). Quand j'ai pris pour la première fois connaissance des résultats de Libet (dans Popper et Eccles, 1977), je l'ai encouragée à y aller voir de plus près, et elle leur a adressé des attaques vigoureuses (Churchland, 1981a). Elle a essayé de discréditer la

première thèse de Libet, la durée longue requise pour atteindre l'« adéquation neuronale » pour la conscience, en demandant aux sujets d'une expérience de dire *go* dès qu'ils seraient conscients de la présence d'un stimulus sur leur peau comparable à ceux utilisés par Libet. Elle a relevé un temps de réponse moyen de 358 msec pour neuf sujets, ce qui, selon elle, montrait que les sujets doivent avoir atteint l'adéquation neuronale vers 200 msec au plus tard (si l'on enlève le temps nécessaire à la production d'une réponse verbale).

La réponse de Libet est stalinienne : une réaction verbale – dire *go* – peut être enclenchée inconsciemment. « Il n'y a rien de magique ou de particulièrement informatif quand la réponse motrice est une vocalisation du mot *go* plutôt que la réponse plus usuelle consistant à appuyer sur un bouton avec son doigt [...] L'aptitude à détecter un stimulus et à réagir à ce stimulus de façon intentionnelle, ou à être psychologiquement influencé par lui, sans pouvoir rapporter une conscience de ce stimulus, est quelque chose de largement accepté » (1981, p. 187-188). À l'objection selon laquelle les sujets de Churchland étaient seulement en train de « dire, comme on le leur demandait, à quel moment précis ils étaient conscients du stimulus », Libet pouvait donner la réponse stalinienne standard : « ils sont finalement devenus conscients du stimulus, mais à ce moment-là, leur compte rendu verbal avait déjà été produit ».

C'est pourquoi Libet rejette des travaux sur les temps de réaction comme ceux de Churchland, parce qu'ils ont, selon lui, « une validité incertaine comme critère principal de l'expérience subjective » (1981, p. 188). Il est partisan de laisser au sujet la possibilité de prendre son temps : « Le compte rendu est fait sans que le sujet se presse en quelques secondes après chaque essai, en permettant au sujet d'examiner introspectivement ce qu'il ressent » (p. 188). Comment, dans ces conditions peut-il prendre en compte la proposition rivale selon laquelle ce temps accordé au sujet donne au révisionniste orwellien dans le cerveau un temps considérable pour remplacer les *vrais* souvenirs par des *faux* ?

« Rapporter ce que l'on éprouve après l'essai requiert bien entendu que les processus de mémoire à court terme et de rappel en mémoire fonctionnent bien, mais cela ne présente aucune difficulté pour les sujets dont les aptitudes ne sont pas significativement défectueuses » (p. 188).

Mais cela revient à commettre une pétition de principe contre l'orwellien, qui est prêt à expliquer toute une gamme d'effets comme des résultats d'un défaut *normal* dans la remémoration ou comme un souvenir hallucinatoire, dans lequel un événement antérieur réel survenu dans la conscience est oblitéré et remplacé par des souvenirs subséquents. Libet a-t-il laissé son ragoût cuire trop longtemps ou bien Churchland l'a-t-elle goûté trop tôt ? Si Libet veut défendre l'idée que son choix du temps de sondage a un statut *privilegié*, il doit être prêt à combattre les contre-arguments.

Libet est prêt à plaider *nolo contendere* : « Il est admis qu'un compte rendu d'ordre temporel relatif ne peut pas, en lui-même, fournir un indicateur du temps "absolu" (d'horloge) de l'expérience : comme il a été suggéré, il n'y a pas de méthode connue pour parvenir à un tel indicateur (1981, p. 188). Cela fait écho à sa remarque antérieure selon laquelle il ne semble y avoir « aucune méthode par laquelle on puisse déterminer l'ordre temporel absolu d'une expérience subjective » (Libet et al., 1979, p. 193). Cependant, Libet néglige que cela peut être dû au fait qu'il n'existe pas de moment de temps absolu de ce genre (cf. Harnad, 1989).

Churchland, dans ses critiques (1981a, 1981b), commet également l'erreur de ne pas distinguer le temps représenté du temps de la représentation :

« Les deux hypothèses diffèrent essentiellement sur la question de savoir quand les deux sensations respectives *ont été ressenties* » (les italiques sont de moi ; 1981a, p. 177).

« Même si l'on suppose que les sensations qui surgissent des sensations simultanées sur la peau et des sensations LM [lemniscus mediant] *sont ressenties exactement au même moment* [les italiques sont de moi], le retard dans l'adéquation neuronale pour les stimuli sur la peau peut aussi bien être un trait créé par le dispositif lui-même » (1981b, p. 494).

Supposez que tous ces artefacts soient éliminés, et que les sensations soient *toujours* « ressenties exactement au même moment ». Comment Churchland interpréterait-elle ce résultat inattendu ? Cela voudrait-il dire qu'il y a un moment *t* tel que le stimulus 1 soit ressenti et que le stimulus 2 soit ressenti en *t* (la perspective antimatérialiste) ou seulement que le stimulus 1 et le stimulus 2 sont ressentis (éprouvés) comme simultanés ?

Churchland ne nous interdit pas d'en inférer que les découvertes de Libet, si elles se confirmaient, ruineraient le matérialisme (comme elle le dit parfois). Ailleurs cependant, elle note à bon droit qu'« aussi intrigantes que puissent être les illusions temporelles, il n'y a pas de raison de supposer qu'il y ait quoi que ce soit de surnaturel en elles, et il n'y a certainement rien qui les distingue d'illusions spatiales ou d'illusions de mouvement, comme si elles seules portaient la marque d'une origine non physique » (1981a, p. 179). Il n'en serait ainsi que si les illusions temporelles étaient des phénomènes dans lesquels *la représentation du temps était systématiquement faussée* ; si les *représentations faussées* avaient lieu aux « mauvais moments », quelque chose de plus révolutionnaire serait en cause.

Que peut-on en conclure au sujet des expériences de Libet sur les stimulations corticales ? Qu'elles sont une tentative intéressante mais peu concluante pour essayer d'établir quelque chose sur *la manière dont le cerveau représente l'ordre temporel*. Les potentiels évoqués primaires peuvent d'une certaine façon servir de points spécifiques de référence pour des représentations neurales du temps, bien que Libet ne l'ait pas montré, comme les critiques techniques de Churchland l'indiquent clairement. D'un autre côté, le cerveau peut rendre ses représentations du temps plus labiles. Nous ne représentons pas des objets vus comme existant sur la rétine, mais plutôt à diverses distances dans le monde extérieur ; pourquoi le cerveau ne représenterait-il pas les événements comme ayant lieu *quand* ils ont le plus de raisons « écologiques » de se produire ? Quand nous sommes engagés dans un acte quelconque de dextérité manuelle, le « temps des doigts » devrait être le critère ; quand nous conduisons un orchestre, le temps « de l'oreille » pourrait saisir ce qui se passe. Le « temps cortical primaire » pourrait être le critère par défaut (un peu comme le temps moyen de Greenwich pour l'Empire britannique). Mais cela mérite d'autres recherches.

Le problème a été embrouillé en raison du fait qu'à la fois partisans et critiques n'ont pas su distinguer de façon cohérente entre temps des représentants et temps des choses représentées. C'est un dialogue de sourds, dans lequel Libet adopte une position stalinienne tandis que Churchland répond de manière orwellienne, tous deux étant apparemment d'accord sur le fait qu'il y a un fait décisif quant au moment exact où (dans le temps

« absolu », comme le dirait Libet) une expérience consciente a lieu ¹⁰.

Les arguments de Libet sur les retards subjectifs d'intentions

Ce concept d'ordre temporel absolu d'une expérience est exploité dans les expériences ultérieures de Libet au sujet des « intentions conscientes ». Dans ces expériences, il a essayé de déterminer cet ordre temporel absolu de façon expérimentale en laissant les sujets, qui seuls ont un accès direct (en un sens) à leurs expériences, faire de l'*automesure* du temps. Il a demandé à des sujets normaux (pas des patients de neurochirurgie) de prendre la « décision spontanée » de fléchir une main tout en notant la position d'un point sur un disque en rotation (la « seconde aiguille » sur une horloge, en fait), au moment précis où ils formaient leur intention (Libet, 1985a, 1987, 1989). Ensuite (quelques secondes après), les sujets rapportaient l'endroit où se trouvait le point au moment où ils avaient décidé de fléchir leur poignet. Cela a permis à Libet de calculer le moment précis (à la milliseconde près) où les sujets *pensaient* qu'ils avaient décidé, et de comparer ce moment avec l'ordre temporel des événements se déroulant concurremment dans leur cerveau. Il a découvert des données établissant que ces « décisions conscientes » s'étaient produites entre 350 et 400 msec derrière le point de départ du « potentiel de préparation » qu'il a été capable d'enregistrer à partir d'électrodes sur le cuir chevelu, qui, selon lui, enregistre les événements neuronaux qui déterminent les actions volontaires accomplies. Il conclut que « la mise en marche cérébrale d'un acte volontaire spontané commence inconsciemment » (1985, p. 259).

Cela semble montrer que votre conscience traîne derrière les processus cérébraux qui contrôlent effectivement votre corps. Beaucoup de gens trouvent que c'est une perspective dérangeante et même déprimante, parce qu'elle semble enlever tout véritable (par opposition à illusoire) « rôle exécutif » au « moi conscient » (voir la discussion des nombreux commentateurs sur Libet, 1985a, 1987, 1989, et dans Harnad, 1982 ; Pagels, 1988, p. 233 *sq.* ; et Calvin, 1989a, p. 80-81).

Libet est plus clair que la plupart de ses critiques quant à

l'importance d'une distinction entre contenu et véhicule : « On ne devrait pas confondre *ce qui* est rapporté par le sujet avec *quand* il peut devenir introspectivement conscient de ce qu'il rapporte » (Libet, 1985a, p. 559).

Il reconnaît (p. 560), de plus, qu'un jugement de simultanéité n'a pas besoin d'être lui-même simultanément produit ou rendu ; il pourrait mûrir pendant une longue période temporelle (considérez, par exemple, les minutes qu'il faut aux commissaires de course pour développer puis examiner la photo finale sur laquelle ils fonderont leur jugement sur le gagnant ou sur le tocard).

Libet a réuni des données sur deux séries temporelles :

– la série objective, qui inclut l'ordre temporel de l'horloge externe et les événements neuronaux saillants : les potentiels de préparation (PP) et les électromyogrammes (EMG) qui enregistrent le début des contractions musculaires ;

– la série subjective (rapportée ensuite) qui consiste en imagerie mentale, en souvenirs de plans préalables, et une seule donnée vérificatrice pour chaque essai : un jugement de simultanéité de la forme : *mon intention consciente (w) a commencé en même temps que le point de l'horloge en position P.*

Libet semble avoir voulu trouver un équivalent approximatif de *l'acte gratuit* dont parlent les existentialistes (par exemple Gide 1948, Sartre 1943), du choix sans motif – et par conséquent, dans certains cas, « libre ». Comme plusieurs commentateurs l'ont fait remarquer, des actions aussi hautement inhabituelles (que l'on peut appeler des actes de pseudo-hasard délibéré) peuvent difficilement passer pour des paradigmes d'« actes volontaires normaux » (Libet, 1987, p. 784). Mais a-t-il isolé une certaine variété d'expériences conscientes, quelle que soit la façon dont on les caractérise, auxquelles on puisse conférer un ordre temporel absolu par ce genre de dispositifs expérimentaux ?

Il soutient que quand on enregistre des intentions conscientes d'agir (au moins des intentions de ce type spécial) en même temps que les événements mentaux qui sont effectivement à l'origine des actes, il y a une durée qui varie dans un intervalle de 30 à 500 msec. C'est énorme – presque une demi-seconde – et cela paraît effrayant à quiconque part du principe que nos actes conscients *contrôlent* nos mouvements corporels. C'est comme si nous étions situés dans des Théâtres Cartésiens dans lesquels on nous montre avec un retard d'une demi-seconde sur la bande du film, le *véritable* processus de décision qui est en train de se

produire *ailleurs* (dans un endroit où *nous* ne sommes pas). Nous ne sommes pas tout à fait « hors du coup » (comme on le dit à la Maison-Blanche), mais puisque notre accès à l'information est ainsi retardé, le mieux que nous puissions faire est d'intervenir en formulant des « vetos » de dernière minute ou des « déclenchements ». Situé en-deçà du Quartier Général (inconscient) de Commandement, je ne peux prendre aucune véritable initiative, je ne suis jamais à la naissance d'un projet, mais j'exerce un peu de modulation exécutrice des politiques déjà formulées qui passent par mon bureau.

Cette image est tentante, mais incohérente. Le modèle de Libet, comme auparavant, est stalinien, mais il existe une alternative orwellienne évidente : les sujets étaient conscients de leurs intentions au moment initial, mais cette conscience a été effacée de leur mémoire (ou seulement révisée) avant qu'ils aient eu une chance de se la rappeler. Libet concède que « cela pose problème, mais que ce n'est pas testable expérimentalement » (1985a, p. 560).

Nonobstant cette concession, la tâche qui consiste à fixer l'ordre temporel absolu de la conscience n'est-elle pas mal conçue ? Ni Libet ni ses critiques ne tirent cette conclusion. Libet, après avoir pris la peine de distinguer le contenu du véhicule — *ce qui* est représenté de *quand* c'est représenté — essaie néanmoins d'inférer des prémisses concernant ce qui est représenté des conclusions concernant l'ordre absolu de la représentation dans la conscience. Le psychologue Gerald Wasserman (1985, p. 556) a bien vu le problème : « On peut déterminer aisément le temps pendant lequel le point externe objectif occupe une certaine position d'horloge, mais ce n'est pas le résultat désiré. » Mais il tombe lui-même dans le piège cartésien : « Ce qui est requis est le temps d'occurrence de la représentation interne du point dans l'esprit-cerveau. »

« Le temps d'occurrence » ou la représentation interne ? Occurrence où ? Une représentation essentiellement continue du point (qui le représente dans diverses positions) est présente dans diverses parties du cerveau, à partir de la rétine et allant vers le système visuel. La luminosité du point est représentée à certains endroits et à certains temps, sa localisation à d'autres, et son mouvement à d'autres encore. Au fur et à mesure que le point extérieur bouge, toutes ces représentations changent, de façon asynchrone et spatialement distribuée. Où « tout se réunit-il en un instant dans la conscience » ? Nulle part.

Wasserman a raison de remarquer que la tâche du sujet consistant à déterminer où se trouvait le point à un certain moment de la séquence subjective est en elle-même une tâche volontaire, et que la mettre en route prend sans doute du temps. C'est difficile non seulement parce que cela entre en compétition avec d'autres projets envisagés au même moment, mais aussi parce que ce n'est pas naturel — c'est un jugement conscient de temporalité d'un type particulier qui ne joue pas normalement de rôle dans le contrôle du comportement, et qui par conséquent n'a pas de signification naturelle dans la séquence. Le processus d'interprétation qui fixe en définitive le jugement de simultanéité subjective est lui-même un artefact de la situation expérimentale et il *change la tâche*, et par conséquent ne nous dit rien d'intéressant sur l'ordre temporel effectif des véhicules normaux de représentation à des endroits quelconques du cerveau.

Il nous faut rejeter une vision trop naturelle : quelque part, au tréfonds du cerveau, l'enclenchement d'un acte commence ; cela commence comme une intention inconsciente, puis il cherche lentement sa voie en direction du Théâtre, gagnant en précision et en pouvoir au fur et à mesure qu'il progresse, et puis, soudain, à l'instant *t*, il surgit sur la scène, où des représentations-de-points sont en marche, qui ont trouvé lentement leur chemin à partir de la rétine, habillées de luminosité et de position spatiale au fur et à mesure qu'elles progressent. L'auditoire (ou plutôt le *Je*) reçoit la tâche de dire quelle est la représentation de point qui est « en scène » exactement quand l'intention consciente a fait son apparition. Une fois identifié, le point de départ de ce point à partir de la rétine peut être calculé, tout comme la distance par rapport au Théâtre et la vitesse de transmission. De cette façon, nous pouvons déterminer le moment exact pendant lequel l'intention consciente est survenue dans le Théâtre Cartésien.

Je trouve extraordinaire le pouvoir de séduction de cette image. Elle est si facile à visualiser ! N'est-ce pas ce qui *doit* se produire quand deux choses se produisent en même temps dans la conscience ? Non. En fait, cela ne *peut pas* être ce qui se passe quand deux choses arrivent en même temps dans la conscience, parce qu'il n'existe pas d'endroit de ce genre dans le cerveau. Certains ont pensé que l'incohérence de *cette* vision n'exigeait pas que nous abandonnions l'idée d'un ordre temporel absolu des expériences. Il existe, semble-t-il, un modèle rival pour le point de

départ de la conscience qui évite l'absurdité du cerveau centralisé de Descartes, tout en permettant un ordre temporel absolu. La conscience ne tient-elle pas non à une arrivée à un point, mais au fait qu'une représentation dépasse un certain seuil d'activation sur tout le cortex ou sur de larges parties de celui-ci ? Selon ce modèle, un élément de contenu devient conscient à un certain temps t , non en entrant dans un système défini fonctionnellement et localisé anatomiquement, mais en changeant d'état à l'endroit même où il se trouve : en prenant une certaine propriété ou en ayant l'intensité de l'une de ses propriétés chauffée au-delà d'un certain seuil.

L'idée que la conscience est un *mode d'action* du cerveau plutôt qu'un sous-système du cerveau est très recommandable (voir par exemple Kinsbourne, 1980 ; Neumann, 1990 ; Crick and Koch, 1990). De plus, les temps de tels changements de mode peuvent être mesurés indépendamment des observateurs extérieurs, fournissant en principe une suite unique et déterminée de contenus atteignant le mode particulier. Mais c'est toujours le Théâtre Cartésien si l'on soutient que l'ordre temporel « absolu » de ces changements de mode définit la séquence subjective. L'imagerie est légèrement différente, mais les implications sont les mêmes. Conférer la propriété spéciale qui caractérise la conscience à un instant ne constitue que la moitié du problème ; déterminer que la propriété a été conférée à cet instant est l'autre moitié, et même si les observateurs scientifiques munis de leurs instruments parviennent à le faire à une échelle microscopique, comment le cerveau le fera-t-il ?

Nous autres êtres humains formons des jugements de simultanéité et serions les éléments de notre propre expérience, dont nous exprimons une partie. Par conséquent, à certains points de notre cerveau, il nous faut franchir la distance entre l'ordre temporel effectif des représentations à la représentation de l'ordre temporel ; partout où et chaque fois que ces discriminations sont effectuées, par la suite, les propriétés temporelles des représentations incorporant ces jugements ne sont pas constitutives de leurs contenus. Les simultanités objectives et les séquences d'événements qui se répandent sur la vaste surface du cortex n'ont pas de pertinence fonctionnelle à moins qu'elles ne puissent être aussi détectées avec précision par des mécanismes dans le cerveau. Nous pouvons formuler une question au sujet du point crucial : en vertu de quoi cette séquence serait-elle le flux de la cons-

cience ? Il n'y a personne à l'intérieur qui soit en train de *regarder* le grand spectacle sur écran montré sur toute la surface du cortex, même si ce spectacle est discernable par des observateurs extérieurs. Ce qui compte, c'est la manière dont ces contenus viennent à être utilisés par ou incorporés dans les processus de contrôle du comportement, et cela ne *doit* être conditionné qu'indirectement par des ordres temporels corticaux. Ce qui compte, une fois de plus, ce ne sont pas les propriétés temporelles des représentants, mais les propriétés temporelles *représentées*, quelque chose qui est déterminé par la manière dont elles sont « prises » par des processus ultérieurs dans le cerveau.

Un bonus :

Le carrousel précognitif de Grey Walter

Après nous être frayé un chemin à travers ces cas complexes, nous méritons de rencontrer quelque chose d'étrange mais de *relativement* facile à comprendre – quelque chose qui nous ramène au message contenu dans ce chapitre difficile. L'expérience de Libet avec l'automesure de temps, comme nous venons de le voir, a créé une tâche délicate de jugement, qui a privé les résultats de la signification escomptée. Le neurochirurgien britannique W. Grey Walter (1963) a réalisé une expérience remarquable qui n'avait pas cet inconvénient. Grey Walter pratiqua son expérience avec des patients dans le cortex moteur desquels il avait implanté des électrodes. Il voulait tester l'hypothèse selon laquelle certains surgissements d'activité enregistrée étaient les initiateurs des actions intentionnelles. Il s'arrangea donc pour que chaque patient regarde des diapositives projetées par un projecteur tournant. Le patient pouvait avancer le carrousel à sa guise, en pressant le bouton de contrôle. (Notez la similarité avec l'expérience de Libet : c'était une décision « libre », rythmée seulement par une montée endogène d'ennui ou de curiosité quant à la prochaine diapositive, par la distraction ou quoi que ce soit d'autre.) À l'insu du patient, cependant, le bouton de contrôle était postiche : il n'était pas attaché au projecteur ! Ce qui faisait avancer les diapositives était en fait le signal amplifié venu de l'électrode implanté dans le cortex moteur du patient.

Nous pourrions supposer que les patients ne remarquaient rien

qui sorte de l'ordinaire. En fait, ils étaient ébahis par l'effet ressenti, parce qu'il leur semblait que le projecteur de diapositives anticipait leurs décisions. Ils rapportaient que juste au moment où ils « s'apprêtaient » à pousser le bouton, mais avant qu'ils aient effectivement décidé de le faire, le projecteur passait les diapositives – et ils se retrouvaient en train de presser le bouton en ayant peur qu'il avance deux fois ! L'effet était saisissant, selon l'analyse de Grey Walter. Mais il semble qu'il n'ait jamais réalisé l'expérience qui aurait été la suite normale de la première : introduire un élément de retard variable pour voir quel retard il fallait incorporer dans le déclenchement pour éliminer l'effet « précognitif » du carrousel.

L'expérience de Grey Walter et celle de Libet sont différentes et cet ordre temporel qui conduit à la surprise dans l'expérience de Grey Walter fait partie d'une tâche normale de contrôle de comportement. De ce point de vue, il ressemble aux jugements d'ordre temporel par lesquels notre cerveau distingue le fait de se mouvoir de gauche à droite du fait de se mouvoir de droite à gauche, plutôt que les jugements « délibérés et conscients » portant sur ces ordres. Le cerveau dans ce cas s'est donné pour tâche d'« attendre » des feedback visuels de l'exécution réussie de son projet d'avancer le carrousel, et le feedback arrive plus tôt que prévu, déclenchant une alarme. Cela devrait nous révéler quelque chose d'important sur l'ordre temporel effectif des véhicules de contenus et des processus qui les accompagnent dans le cerveau, mais cela ne nous montrerait, contrairement aux apparences, rien quant à « l'ordre temporel absolu de la décision consciente de changer la diapositive ».

Supposez, par exemple, qu'une extension de l'expérience de Grey Walter montre qu'un retard allant jusqu'à 300 msec (comme l'implique l'expérience de Libet) doive être incorporé dans la réalisation de l'acte afin d'éliminer le sentiment subjectif d'un changement de diapositives précognitif. Ce que montrerait en fait l'existence d'un tel retard serait que les attentes qui sont initiées par une décision de changer la diapositive s'accordent avec l'attente d'un feedback visuel 300 msec plus tard, et avec l'envoi en retour d'une alarme dans d'autres conditions. (C'est l'analogie d'un message exprimant un choix venu du commandant en chef à Calcutta envoyé à Whitehall au moment du début de la bataille de La Nouvelle-Orléans.) Le fait que l'alarme soit *finale*ment interprétée dans la séquence subjective comme une

perception d'événements désordonnés (changement avant de pousser le bouton) ne montre rien quant au *moment* où en temps réel, la conscience de la décision a surgi en premier lieu. Le sentiment que rapportent les sujets de ne pas avoir eu tout à fait le temps d'opposer un « veto » à la pression initiale du bouton quand « ils ont vu que la diapositive était déjà en train de changer » est une interprétation naturelle que peut produire le cerveau sur les contenus variés qui sont finalement disponibles à différents moments pour être incorporés dans le récit. Ce sentiment était-il déjà présent au premier moment de conscience de l'intention (auquel cas l'effet requiert un long délai pour « montrer le temps » et il est stalinesque), ou bien était-ce une interprétation rétrospective et un *fait accompli* qui prête par ailleurs à confusion (auquel cas il est orwellien) ? J'espère qu'à présent il est clair que les présuppositions qui affectent cette question la disqualifient.

Derniers détails à régler

Vous avez peut-être encore envie d'objecter que tous les arguments avancés dans ce chapitre ne parviennent pas à repousser la vérité évidente selon laquelle nos expériences des événements surviennent dans le même ordre que celui dans lequel nous faisons l'expérience qu'ils surviennent. Si quelqu'un pense la pensée « un, deux, trois, quatre, cinq », sa pensée au sujet de « un » survient avant sa pensée au sujet de « deux », et ainsi de suite. Cet exemple illustre une thèse qui est en général correcte et qui semble bien entendu ne pas souffrir d'exceptions tant que nous limitons notre attention à des phénomènes psychologiques de durée « ordinaire » et macroscopique. Mais les expériences que nous avons examinées portent sur des événements qui sont confinés à des cadres temporels inhabituellement étroits et durent quelques centaines de millisecondes. À cette échelle, l'hypothèse standard s'effondre.

Tout événement dans votre cerveau a une localisation spatio-temporelle définie. Mais demander quand vous êtes devenu conscient du stimulus suppose que l'un de ces événements est, ou revient à être, devenu conscient du stimulus. C'est comme demander quand exactement l'Empire britannique a été informé du traité dans la guerre de 1812. À un certain moment, entre

le 24 décembre 1814 et la mi-janvier 1815. C'est assez précis, mais il ne s'est pas produit de fait décisif qui puisse être déterminé au jour et à l'heure près. Même si nous pouvons donner des temps précis pour les divers moments où les dirigeants de l'Empire ont été informés, aucun de ces moments ne peut être considéré comme celui où l'Empire lui-même est devenu informé. La signature du traité lui-même était un acte officiel, intentionnel de l'Empire, mais la participation des troupes britanniques à la bataille de La Nouvelle-Orléans en était un autre, et il a été accompli selon l'hypothèse où aucun traité n'avait été signé. On peut prétendre défendre l'idée que l'arrivée de la nouvelle à Whitehall ou à Buckingham Palace à Londres doit être considérée comme le temps officiel où l'Empire a été informé, puisque c'était le « centre nerveux » de l'Empire. Descartes pensait que la glande pinéale était précisément un tel centre nerveux dans le cerveau, mais il avait tort. Puisque la cognition et le contrôle – et par conséquent la conscience – sont distribués à travers le cerveau, aucun moment ne peut compter comme le moment précis où se produit chaque événement conscient.

Dans ce chapitre, j'ai essayé d'ébranler certaines mauvaises habitudes de pensée et de les débouter de leurs « fondements » arbitraires pour les remplacer par de meilleures façons de penser. Mais en chemin j'ai été obligé d'abandonner bien des impasses. La plus tentante d'entre elles, je crois, est l'assertion métaphorique selon laquelle le « sondage » est ce qui « précipite les récits ». L'ordre temporel des sondages d'inspection effectués par les expérimentateurs peut, selon moi, avoir un effet majeur de révision sur les systèmes de représentation qu'utilise le cerveau. Mais parmi ceux qui peuvent diriger ce genre de sondages d'inspection sur le sujet se trouve le sujet lui-même. Si vous vous intéressez à la question de savoir précisément quand vous êtes devenu conscient de quelque chose, votre propre examen de vous-même ou vos auto-inspections fixent des limites pour de nouvelles séquences de contrôle, altérant ainsi les conditions des processus impliqués.

Les résultats des sondages effectués par des observateurs extérieurs sont typiquement des actes de langage d'un type ou d'un autre, et ces derniers *expriment des jugements* sur divers contenus de conscience, que chacun peut entendre et interpréter. Les résultats des autosondages sont des entités *de la même catégorie sémantique*. Ce ne sont pas des « présentations » (dans le Théâtre

Cartésien) mais des *jugements* sur ce qui apparaît au sujet, des jugements que le sujet lui-même peut interpréter, qu'il peut utiliser pour ces actions et dont il peut se souvenir. Dans les deux cas, ces événements fixent des interprétations de ce dont le sujet a fait l'expérience ; ils fournissent ainsi des points fixes dans la séquence subjective. Mais, selon le modèle des Versions Multiples, il n'y a pas *en plus* d'un tel jugement, et des discriminations antérieures sur lesquelles il est fondé, de *présentation* du matériau-à-interpréter à laquelle puisse accéder le Juge en Chef, l'auditoire du Théâtre Cartésien. C'est une idée qui n'est toujours pas facile à comprendre et même à admettre. Nous devons parvenir à cette idée par plusieurs voies.

NOTES

1. Les arguments et les analyses de ce chapitre (et certaines des discussions du chapitre précédent) proviennent des travaux présentés dans Dennett et Kinsbourne (1992).

2. Cela ne veut pas dire que le cerveau n'utilise jamais de « souvenirs tampons » pour servir d'interface entre ses processus internes et le monde extérieur asynchrone. La « mémoire tampon » avec laquelle nous préservons brièvement des trames de stimuli pendant que le cerveau commence à les traiter est un exemple clair de ce phénomène (Sperling, 1960 ; Neisser, 1967 ; voir aussi Newell, Rosenbloom, et Laird 1989, p. 107).

3. Je m'empresse d'ajouter que je m'autorise cet embellissement historique. Francis Rawdon-Hastings, le premier marquis d'Hastings et second comte de Moira, était gouverneur général du Bengale et commandant en chef en Inde en 1815, mais je n'ai pas la moindre idée de quand ni de comment il fut informé de la bataille de La Nouvelle-Orléans.

4. Un tel « tampon postal » peut en principe être ajouté à un véhicule de contenu à n'importe quelle étape de son voyage ; si tous les messages qui arrivent à une certaine destination viennent du même lieu, par le même chemin et à la même vitesse, leur « date de départ » à partir de la destination originale peut être rétroactivement tamponnée sur eux, en soustrayant simplement une constante de leur temps d'arrivée à la gare. Cette possibilité d'ingénierie est sans doute utilisée par le cerveau pour faire certains ajustements automatiques pour des temps de voyage standards.

5. Comme le note Uttal (1979), cette distinction *est* largement reconnue par les chercheurs en neurosciences : « L'essence de la majeure partie de la recherche qui a été effectuée dans le domaine du codage sensoriel peut se résumer à une idée unique et très importante – tout code possible peut représenter n'importe quelle dimension spatiale ; on n'a pas besoin d'une relation isomorphe entre les donnés neurales et les donnés psycho-physiques. L'espace peut représenter le temps, le temps peut représenter des fonctions psycho-physiques linéaires ou non linéaires aussi bien » (p. 286). Mais bien que cette idée soit bien connue, nous allons bientôt voir que certains chercheurs la comprennent en se *trompant* sur son sens ; ils la comprennent en réintroduisant tacitement l'« isomorphisme » inutile dans un processus obscur de traduction ou de « projection » supposé se produire dans la conscience.

6. Cf. Pylyshyn (1979, p. 278) : « Personne [...] n'est disposé à parler *au sens*

littéral de propriétés physiques d'un événement mental telles que sa couleur, sa taille, sa masse, et ainsi de suite... bien que nous parlions en fait de ces propriétés comme *représentant* (ou ayant le contenu représentationnel de) telles propriétés. Par exemple, on ne dirait pas à proprement parler d'une pensée (image) qu'elle est grande ou verte, mais seulement qu'elle portait *sur* quelque chose de grand ou de vert (ou que c'était une image *de* quelque chose de grand ou de vert)... Nous devrions par conséquent trouver curieux que l'on appelle librement de la *durée* d'un événement mental.

7. Comme le remarque le psychologue Robert Efron : « Quand nous observons en premier un objet avec la vision centrale, nous ne faisons pas vaguement l'expérience de l'objet comme il apparaîtrait avec une vision moins périphérique... De même, quand nous déplaçons notre attention d'un objet de conscience à un autre, il n'y a pas d'expérience d'une spécificité "accrue" des nouveaux objets venus à la conscience. Nous percevons simplement le nouvel objet » (1967, p. 721).

8. Voyez aussi comment Libet rejette l'interprétation plus modérée suggérée par Mc Kay (1981, p. 195 ; 1985b, p. 568). Le dernier mot de Libet sur ces sujets en 1981 ne contient pas de conclusion explicite : « Ma thèse... a été que le contraste temporel crée des difficultés relatives pour la théorie de l'identité, mais que celles-ci ne sont pas insurmontables » (p. 196). On peut présumer qu'elles seraient sans aucun doute insurmontables selon l'interprétation de la *projection* à rebours, parce que cela implique une pré cognition, ou une causalité à rebours ou quelque chose d'au moins bizarre et de jamais vu. De plus Libet décrit plus tard (1985, p. 569) ces difficultés non insurmontables de telle façon que l'interprétation modérée semble ici requise : « Bien que l'hypothèse du retard et de l'antidatage ne sépare pas le temps effectif de l'expérience du temps de la production neuronale, elle n'élimine pas la nécessité d'une simultanéité entre l'ordre temporel *subjectif* de l'expérience et le temps effectif objectif de l'expérience. » Peut-être le soutien enthousiaste de Sir John Eccles à une interprétation radicale, dualiste des découvertes, a-t-il détourné l'attention de Libet et de ses critiques de la thèse modérée qu'il soutient lui-même.

9. Dans un article antérieur, Libet concédait la possibilité de processus orwelliens et supposait qu'il pourrait y avoir une différence significative entre des événements mentaux inconscients et des événements mentaux conscients, mais éphémères : « Il pourrait y avoir un type d'expérience immédiate bien qu'éphémère de conscience mais qui ne soit pas retenue pour être mémorisée à des niveaux conscients de l'expérience. Mais, si de telles expériences existent, leurs contenus auraient une signification directe seulement dans des processus inconscients ultérieurs, bien que, comme d'autres expériences conscientes, elles pourraient jouer un rôle indirect dans des expériences conscientes ultérieures » (1965, p. 78).

10. Harnad (1989) voit un problème insoluble de mesure, mais nie précisément ce que je soutiens – à savoir qu'il n'y a pas de fait décisif : « L'introspection peut seulement nous dire quand un événement a *semblé* survenir, ou lequel de deux événements a *semblé* survenir en premier. Il n'y a pas de façon indépendante de confirmer que le véritable ordre temporel était en fait celui qui semblait être tel ou tel. L'incommensurabilité est un problème méthodologique, pas un problème métaphysique » (p. 183).

Chapitre 6

L'évolution de la conscience

« Toute chose est ce qu'elle est parce qu'elle est devenue telle. »

D'Arcy Thomson, 1917.

À l'intérieur de la boîte noire de la conscience

La théorie esquissée au chapitre précédent commence à montrer comment la conscience pourrait résider dans le cerveau humain. Sa contribution principale était négative : elle visait à ébranler la dictature exercée par la conception du Théâtre Cartésien. Nous avons *commencé* à remplacer ce modèle par une autre image positive, mais nous ne sommes pas allés bien loin. Pour progresser encore, nous devons changer de terrain et aborder les complexités de la conscience d'un point de vue différent : celui de l'évolution. Comme la conscience humaine n'a pas toujours existé, elle a dû naître de phénomènes antérieurs qui n'étaient pas eux-mêmes conscients. Si nous nous penchons sur les phénomènes qui doivent avoir – ou pourraient avoir – été impliqués dans cette transition, peut-être parviendrons-nous à mieux considérer ces phénomènes complexes et leur rôle dans la création du phénomène complet de la conscience.

Dans son élégant petit livre intitulé *Vehicles : Essays in Synthetic Biology* (1984), le neurobiologiste Valentino Braitenberg décrit une série de mécanismes autonomes de plus en plus compliqués ; à partir de dispositifs d'une simplicité dérisoire et totalement dénués de vie, il a élaboré petit à petit des entités

(imaginaires) qui ont une apparence de vie et d'activité psychologique étonnante. Cet exercice d'imagination marche en raison de ce qu'il appelle la loi d'analyse ascendante et de synthèse descendante : il est plus facile d'imaginer le comportement (et les implications comportementales) d'un dispositif que l'on synthétise « du dedans au dehors », pour ainsi dire, que d'essayer d'analyser le comportement extérieur d'une « boîte noire » et de se représenter ce qui doit se produire à l'intérieur.

Pour le moment, nous avons en fait traité la conscience elle-même comme une boîte noire de ce genre. Nous avons pris son « comportement » (c'est-à-dire sa phénoménologie) comme un « donné » et nous nous sommes demandés quelles sortes de mécanismes cachés dans le cerveau pourraient l'expliquer. Renversons à présent notre stratégie, et envisageons l'évolution des mécanismes cérébraux qui permettent d'accomplir telle ou telle chose, et voyons si ce qui en sort nous fournit une hypothèse plausible du mécanisme qui expliquerait certains des « comportements » problématiques de nos cerveaux conscients.

Beaucoup de théories – en fait, beaucoup de spéculations – ont cherché à expliquer l'évolution de la conscience humaine, à commencer par les remarques de Darwin lui-même dans *La Descendance de l'homme* (1871). À la différence de la plupart des explications en sciences, les explications évolutionnistes sont essentiellement des récits, qui nous ramènent d'une époque où quelque chose n'existait pas à une époque où quelque chose existe par une série d'étapes que le récit a pour charge d'expliquer. Plutôt que d'essayer de passer en revue de façon érudite tous les récits qu'on a pu inventer, je me propose de raconter une seule histoire, en empruntant librement mon matériau à de nombreux autres théoriciens et en me concentrant sur quelques points qu'on a souvent négligés ; ils nous aideront à lever certains obstacles que nous rencontrons quand nous essayons de comprendre la nature de la conscience. J'ai résisté à la tentation d'inclure littéralement des douzaines d'épisodes secondaires, et quand j'avance ou rejette une thèse, je me suis dispensé de donner tous les arguments pour ou contre, comme sont enclins à le faire la plupart des philosophes. Le résultat, je le reconnais, ressemble un peu à un résumé en cent mots de *Guerre et paix*. Mais nous n'avons pas que cela à faire ¹.

L'histoire que nous devons raconter ressemble à d'autres histoires que la biologie a commencé à nous raconter. On peut la

comparer, par exemple, à l'histoire des origines de la sexualité. Il existe aujourd'hui beaucoup d'organismes qui ne sont pas sexués et qui se reproduisent asexuellement ; il fut un temps où aucun des organismes existants n'avait de sexe, mâle ou femelle. En quelque façon, par l'effet d'une série d'étapes que nous pouvons nous représenter, certains de ces organismes doivent avoir évolué de manière à devenir des organismes sexués, pour produire, en définitive, quelque chose comme nous-mêmes. Quelles étaient les conditions requises pour stimuler ou rendre nécessaires ces innovations ? Pourquoi, en bref, tous ces changements se sont-ils produits ? Ce sont quelques-uns des problèmes les plus profonds qui se posent à la théorie évolutionniste contemporaine ².

La question de l'origine de la sexualité et celle des origines de la conscience sont parallèles. La vie sexuelle des fleurs, des huîtres, et d'autres formes simples de vie n'a presque rien d'*érotique* (en termes humains), mais nous pouvons reconnaître dans leurs routines de reproduction, mécaniques et apparemment sans joie, les bases et les principes de notre monde sexuel, lequel est autrement plus excitant. De même, les précurseurs primitifs des mois humains conscients n'ont rien de particulièrement *égoïque* (si je peux forger ce terme), mais ils jettent les bases de nos innovations et complications humaines particulières. La conception de notre esprit conscient est le résultat de trois processus évolutifs successifs, empilés les uns sur les autres, chacun étant plus rapide et plus puissant que son prédécesseur. Pour comprendre cette pyramide de processus, nous devons commencer au commencement.

Les débuts

Scène une : la naissance des frontières et des raisons

Au début, il n'y avait pas de raisons ; il n'y avait que des causes. Rien n'avait un but, et rien n'avait quoi que ce soit qui puisse ressembler à une fonction ; il n'y avait pas du tout de téléologie dans le monde. L'explication en est simple : il n'y avait pas d'être doué d'intérêts. Mais après des millénaires apparurent des *réplicateurs* simples (R. Dawkins, 1976 ; voir aussi Monod,

1970, chapitre 1, « D'étranges objets »). Alors qu'ils n'avaient pas la moindre idée de leurs intérêts, et peut-être à proprement parler pas d'intérêts, nous pouvons, en scrutant rétrospectivement l'horizon à partir de notre point de vue divin sur leurs débuts, leur assigner de manière arbitraire certains intérêts, engendrés par leur « intérêt » spécifique à se reproduire. Cela ne faisait peut-être pas de différence, n'impliquait aucun objectif; cela n'importait à qui que ce soit ou à quoi que ce soit qu'ils aient ou non réussi à se répliquer (bien qu'il semble que nous puissions leur être reconnaissants de l'avoir fait), mais au moins nous pouvons leur assigner des intérêts sous certaines conditions. Si ces réplicateurs simples doivent survivre et se reproduire, persistant ainsi face à une entropie croissante, leur environnement doit satisfaire certaines conditions : des conditions propres à produire la réplication doivent être présentes ou au moins fréquentes.

En termes plus anthropomorphiques, si ces réplicateurs simples veulent continuer à se répliquer, ils doivent espérer diverses choses et tendre vers elles : ils doivent éviter les choses « mauvaises » et rechercher celles qui sont « bonnes ». Quand une entité arrive sur la scène et qu'elle est capable d'un comportement qui prévienne, de manière plus ou moins primitive, sa propre dissolution et sa propre décomposition, elle introduit avec elle dans le monde son « bien ». En d'autres termes, elle crée un point de vue à partir duquel les événements du monde peuvent être découpés grossièrement en événements favorables, défavorables ou neutres. Et ses propres tendances à rechercher les premiers, à éviter les seconds et à ignorer les troisièmes types d'événements contribuent de façon essentielle à la définition de ces trois classes. Au fur et à mesure que la créature vient à avoir des intérêts, le monde et ses événements commencent à créer des *raisons* pour cela — même si la créature ne les reconnaît pas complètement ou pas du tout (Dennett, 1984a). Les premières raisons ont pré-existé à leur propre reconnaissance. En fait, le premier problème qu'ont rencontré les êtres capables de se trouver confrontés à des problèmes a été d'apprendre comment reconnaître et agir sur la base des raisons que leur existence même a fait exister.

À partir du moment où quelque chose se préoccupe de se préserver soi-même, les frontières deviennent importantes. En effet, si l'on se met à se préserver soi-même, on ne veut pas gaspiller ses efforts à essayer de préserver le monde tout entier :

on trace une limite. On devient, en un mot, *égoïste*. Cette forme primordiale d'égoïsme (qui, en tant que forme primordiale, n'a pas la plupart des caractéristiques de *notre* propre type d'égoïsme) est l'un des signes distinctifs de la vie. L'endroit où un morceau de granit finit et où un autre commence est très ténu : la frontière de la fracture peut être réelle, mais rien ne travaille à protéger le territoire, à repousser la frontière ou à exiger un recul. « Moi et le monde contre moi » – cette distinction entre tout ce qui se passe à l'intérieur d'une limite fermée et tout ce qui se passe dans le monde extérieur – est au cœur de tous les processus biologiques, pas seulement ceux de l'ingestion et de l'excrétion, de la respiration et de la transpiration. Prenons, par exemple, le système immunitaire, avec ses millions d'anticorps différents disposés en ligne de défense du corps contre des millions d'envahisseurs étrangers. Cette armée doit résoudre le problème fondamental de la reconnaissance : distinguer son soi (et celui de ses amis) de tout le reste. Le problème a été résolu d'une façon assez proche de celle par laquelle les nations humaines, et leurs armées, ont résolu le problème homologue : par l'intermédiaire de techniques routinières d'identification – les officiers des douanes et des passeports en miniature sont des formes moléculaires et des détecteurs de formes. Il est important de reconnaître que cette armée d'anticorps n'a pas de généraux, pas de QG ni de plan de bataille, ni même de description de l'ennemi : les anticorps représentent leurs ennemis seulement à la façon dont un million de serrures représentent les clefs qui les ouvrent.

Il nous faut relever plusieurs autres faits qui sont déjà évidents à ce stade initial. Alors que l'évolution dépend de l'histoire, Dame Nature n'est pas snob, elle ne prend pas garde aux origines. Peu importe si ou comment un organisme a fait des prouesses : il vaut ce qu'il fait. Autant que nous le sachions, les pedigrees des répliqueurs initiaux étaient tous à peu près semblables : ils étaient tous les produits d'une série aveugle, sourde et aléatoire de sélections quelconques. Mais si un hyper-ingénieur voyageant dans le temps avait inséré un *répliqueur-robot* dans tout cela, et si ses prouesses avaient été égales ou meilleures par rapport à celles de ses rivaux naturels, ses descendants pourraient bien être parmi nous – ils pourraient même *être* nous ! (Dennett, 1987a, 1990b).

La sélection naturelle ne peut pas dire comment un système est devenu ce qu'il est devenu, mais cela ne veut pas dire que des différences profondes ne pourraient pas séparer des systèmes

« conçus » par la sélection naturelle et ceux qui sont conçus par des ingénieurs humains intelligents (Langton, Hogeweg, in Langton, 1989). Par exemple, les concepteurs humains, qui voient loin mais qui portent des œillères, tendent à découvrir que leurs conceptions sont déjouées par des effets secondaires et des interactions inattendues ; c'est pourquoi ils essaient de se prévenir contre ces effets en donnant à chaque élément du système une fonction unique, et en l'isolant des autres éléments. En revanche, Dame Nature (le processus de la sélection naturelle) est bien connue pour être myope et pour ne pas avoir de buts. Puisqu'elle ne prévoit rien du tout, elle n'a aucun moyen de s'inquiéter des effets secondaires qu'elle ne prévoit pas. En n'« essayant » pas de les éviter, elle tente des conceptions dans lesquelles interviennent de nombreux effets secondaires ; la plupart de ces conceptions sont épouvantablement défectueuses (demandez à un ingénieur), mais à tout moment un *effet secondaire collatéral inattendu* apparaît ; deux systèmes fonctionnels ou plus qui ne sont pas liés l'un à l'autre peuvent interagir de manière à produire un bonus : des fonctions multiples pour des éléments simples. Les fonctions multiples ne sont pas inconnues dans les artefacts des ingénieurs humains, mais elles sont relativement rares ; dans la nature elles sont partout ; et, comme nous le verrons, si les théoriciens ont eu tellement de mal à trouver des conceptions plausibles pour la conscience dans le cerveau, c'est parce qu'ils ont tendu à penser les éléments du cerveau comme ayant chacun une fonction spécifique ³.

C'est ainsi qu'on peut poser les fondations. Nous pouvons à présent expliquer les faits plus fondamentaux primordiaux :

1. Il y a des raisons à reconnaître.
2. Là où il y a des raisons, il y a des points de vue à partir desquels on peut les reconnaître et les organiser.
3. Tout agent doit distinguer « ici dedans » du « monde extérieur ».
4. Toute reconnaissance doit en dernière instance être accomplie par une myriade de routines « aveugles et mécaniques ».
5. À l'intérieur de la limite défendue, il n'y a pas nécessairement un Officier Supérieur ou un Quartier Général.
6. Dans la nature, on vaut ce que l'on fait ; les origines ne comptent pas.

7. Dans la nature, les éléments ont souvent des fonctions multiples dans l'économie d'un organisme simple.

Nous avons déjà pu percevoir des échos de ces faits primordiaux dans la recherche d'un « point de vue ultime de l'observateur conscient » et dans plusieurs cas où nous avons remplacé les homoncules par des (équipes de) mécanismes simples. Mais, comme nous l'avons vu, le point de vue d'un observateur conscient n'est pas identique aux points de vue primordiaux des premiers répliqueurs qui ont divisé leur monde en bon et mauvais, mais il est un descendant sophistiqué de ces points de vue. (Après tout, en ce sens primordial, même les plantes ont des points de vue.)

Scène deux :

Comment produire le futur mieux et à nouveaux frais

« Et l'une des fonctions les plus profondes, les plus générales des organismes vivants est d'aller de l'avant, de produire le futur, comme l'a dit Paul Valéry. »

François Jacob, 1982, p. 66.

« Prédire le futur d'une courbe c'est effectuer une certaine opération sur son passé. On ne peut pas réaliser la vraie opération de prédiction par un appareillage que nous puissions construire, mais il y a certaines opérations qui lui ressemblent, et qu'on peut en fait réaliser par un appareillage que nous puissions construire. »

Norbert Wiener, 1948, p. 12.

Dans le chapitre précédent, j'ai mentionné en passant que le but principal des cerveaux était de produire le futur. C'est une idée qui mérite qu'on s'y attarde un peu plus. Pour s'en tirer, un organisme doit ou bien se cuirasser lui-même (comme un arbre ou une palourde) et « espérer que tout ira pour le mieux », ou bien développer des méthodes pour se prémunir des obstacles et chercher les meilleurs voisinages dans son espace proche. Si l'on suit la seconde méthode, on est confronté au problème principal que doit résoudre tout *agent* : que faire maintenant ?

Pour résoudre ce problème, il vous faut un système nerveux, afin de *contrôler* vos activités dans le temps et dans l'espace. La

jeune ascidie vagabonde à travers la mer à la recherche d'un rocher ou d'un massif de corail bien choisi auquel elle pourrait s'accrocher pour établir sa demeure à vie. Pour cela, elle dispose d'un système nerveux rudimentaire. Quand elle trouve son lieu d'élection et prend racine, elle n'a plus besoin de son cerveau ; c'est pourquoi elle le mange ! (C'est un peu comme quand un enseignant est titularisé ⁴.) La clef du contrôle, c'est l'aptitude à *suivre à la trace* ou même à *anticiper* les traits importants de l'environnement, si bien que tous les cerveaux sont, par essence des *machines à anticipation*. La coque de la palourde est une armure fine, mais elle ne peut se tenir toujours fermée ; le réflexe câblé qui provoque la fermeture de la coque est un anticipateur/éviteur de dangers fruste mais efficace.

Plus primitives encore sont les réponses d'approche et de retrait des organismes les plus simples, qui sont liées de la façon la plus directe et la plus inévitable aux sources de ce qui est bénéfique et de ce qui est mauvais : les organismes *touchent* ces dernières. Puis, selon que la chose touchée est mauvaise ou bonne pour eux, ils reculent ou plongent (dans l'instant, s'ils sont chanceux). Ils font cela en étant simplement « câblés » de telle sorte que le contact avec le trait bon ou mauvais dans le monde déclenche le tortillement opportun. Ce fait est la base, comme nous le verrons, de certaines des propriétés les pires et les plus délicieuses (littéralement) de la conscience. Au début, *tous* les « signaux » causés par des choses de l'environnement signifiaient soit « casse-toi ! », soit « vas-y ! » (Humphrey, à paraître).

Aucun système nerveux, à ses débuts, n'avait la possibilité d'utiliser un « message » plus dépassionné et plus objectif l'*informant* simplement, de façon neutre, sur telle ou telle circonstance. Mais des systèmes nerveux aussi simples ne peuvent pas tirer grand-chose du monde. Ils sont capables seulement de ce que nous pouvons appeler des anticipations *proximales*, comportement adapté à ce qui se produit dans le futur *immédiat*. Les cerveaux plus évolués, au contraire, peuvent extraire plus d'information et plus rapidement ; ils peuvent l'utiliser pour *éviter* le contact nuisible au premier chef ou pour *chercher* les éléments nutritifs (les opportunités de compagnonnage, à partir du moment où le sexe fit son apparition).

Confrontés à la tâche consistant à extraire du futur utile de notre passé personnel, des organismes comme le nôtre essaient d'obtenir quelque chose gratuitement (ou tout au moins à prix

réduit) : trouver les lois du monde – et s'il n'y en a pas, trouver les lois approximatives du monde –, c'est-à-dire tout ce qui pourra nous donner une limite. De certains points de vue, il apparaît extrêmement remarquable que des organismes comme nous puissent tirer quoi que ce soit de la nature. Pour quelle raison profonde la nature nous tuyaute-t-elle ou révèle-t-elle ses régularités quand on vient à l'inspecter ? Tout dispositif qui produit du futur utile est susceptible d'être quelque chose comme une technique – un système de fortune qui se trouve fonctionner, le plus souvent possible, et atteindre par chance une régularité du monde que l'on peut détecter. Les anticipateurs chanceux sur lesquels Dame Nature peut tomber doivent être prisés, bien entendu, s'ils améliorent les limites d'un organisme.

Au plus bas de l'échelle, par conséquent, nous trouvons les créatures qui représentent le moins possible : juste assez pour *laisser le monde les avertir* quelquefois quand ils font quelque chose qui ne va pas. Les créatures qui suivent cette politique ne s'engagent pas dans des activités de planification. Elles plongent en avant, et si quelque chose commence à leur faire mal, ils en « savent assez » pour abandonner, mais c'est le mieux qu'ils puissent faire.

L'étape suivante implique une anticipation à court terme – par exemple la capacité à esquiver les obstacles. Ce talent anticipatoire est souvent « câblé » – il fait partie de la machinerie innée conçue sur des millions d'années pour détecter les régularités (sujettes à des exceptions) que nous pouvons relever entre les *choses qui menacent* et les *choses qui nous touchent*. La réponse d'évitement des objets qui passent est câblée profondément chez les êtres humains, par exemple, et on peut l'observer chez les nouveau-nés (Yonas, 1981) : c'est un don de nos lointains ancêtres, alors que leurs cousins qui n'ont pas survécu ne savaient pas suffisamment éviter les obstacles. Est-ce que le signal « quelque chose approche » *signifie* « tirons-nous ! » ? Disons qu'il proto-signifie cela ; il est câblé directement au mécanisme d'évitement.

Nous avons d'autres dons du même type. Nos systèmes visuels, comme ceux de beaucoup d'autres animaux, même les poissons, sont très sensibles à des structures dotées d'un axe de symétrie verticale. Selon Braitenberg, c'est probablement parce que dans le monde naturel de nos lointains ancêtres (bien avant qu'il n'y ait des façades d'églises ou des ponts suspendus), à peu près les seules choses dans le monde qui *manifestaient* des axes de symé-

trie verticale étaient d'autres animaux, et seulement quand ils vous faisaient face. C'est pourquoi nos ancêtres ont été équipés d'un système d'alarme très efficace se déclenchant (la plupart du temps) quand ils étaient *contemplés* par un autre animal (Braitenberg, 1984) ⁵. Identifier un prédateur à une certaine distance (spatiale) au lieu de devoir attendre le moment où il plante ses dents dans votre chair est aussi une sorte d'anticipation *temporellement* plus distale : elle vous donne une tête d'avance pour éviter l'obstacle.

Ces mécanismes caractérisent aussi par leur simplicité discriminatoire : ils échangent ce que l'on pourrait appeler *la vérité et la précision des rapports* contre la *vitesse et l'économie*. Certaines choses qui déclenchent un détecteur de symétrie verticale n'ont pas de signification effective pour l'organisme : l'aspect quasi symétrique peu fréquent d'un arbre ou d'un arbuste ou (à l'époque moderne) de beaucoup d'artefacts humains. Par conséquent, la classe des choses que le mécanisme distingue forme, officiellement, une foule désordonnée — dominée par les animaux-qui-regardent-dans-ma-direction, mais permettant une variété indéfiniment grande de fausses alarmes (relativement à *ce* message). Et ce ne sont même pas toutes les structures dotées de symétrie verticale ou seulement elles mettent en marche le mécanisme ; certaines structures symétriques verticales ne parviendront pas, pour une raison ou pour une autre, à le déclencher, et il y aura de fausses alarmes dans ce cas également ; c'est le prix à payer pour avoir un mécanisme rapide, bon marché, et portable ; et c'est un prix que les organismes, dans leur narcissisme (Akins, 1989) sont prêts à payer. Il est facile de constater ce fait, mais certaines de ses implications pour la conscience ne sont pas si évidentes de prime abord. (Au chapitre 11, cela deviendra important quand nous poserons des questions comme : quelles propriétés détectons-nous avec notre vision des couleurs ? Qu'est-ce que les choses rouges ont en commun ? Et même : pourquoi le monde nous apparaît-il de la façon dont il nous apparaît ?)

Être informé avec une certaine marge d'erreur de ce qu'un autre animal vous regarde est presque toujours un événement qui a un sens dans le monde naturel. Si l'animal ne veut pas vous manger, il peut être un compagnon potentiel, ou un rival dans la quête d'un compagnon, ou une proie qui a décelé votre approche. L'alarme doit enclencher ensuite les analyseurs « ami, ennemi ou nourriture ? », en sorte que l'organisme puisse dis-

tinguer des messages tels que « Un membre de mon espèce est en train de me regarder ! », « Un prédateur vous a dans sa ligne de mire ! » et « Votre souper est prêt à se tirer ! ». Chez certaines espèces (certains poissons, par exemple), le détecteur de symétrie verticale est câblé de manière à produire une interruption rapide de l'activité présente connue comme *réponse d'orientation*.

Selon le psychologue Odmar Neumann (1990), les réponses d'orientation sont l'homologue biologique de l'alarme à bord d'un bateau : « Tout le monde sur le pont ! » La plupart des animaux, comme nous, ont des activités qu'ils contrôlent de façon routinière, « en pilotage automatique » ; ils n'utilisent pas toutes leurs capacités, mais sont sous le contrôle de sous-systèmes spécialisés de leurs cerveaux. Quand une alarme spécialisée se déclenche (comme l'alarme pour l'objet qui passe ou l'alarme de la symétrie verticale chez nous) ou quand une alarme générale se déclenche sous l'effet de quelque chose de soudain et de surprenant (ou de simplement inattendu), le système nerveux de l'animal se mobilise pour prendre en compte la possibilité d'une urgence. L'animal arrête ce qu'il est en train de faire et balaie rapidement l'espace environnant ou fait une mise à jour qui donne à tout organe sensoriel la possibilité de contribuer à la masse des informations pertinentes disponibles. Une aire de contrôle *temporellement* centralisée s'établit à travers une activité neuronale déterminée — toutes les lignes sont ouvertes, pour une brève période. Si le résultat de cette mise à jour est le déclenchement d'une « seconde alarme », le corps tout entier de l'animal se mobilise, par une décharge d'adrénaline. Sinon, l'activité déterminée diminue rapidement, l'équipe rendue inactive retourne au lit, et les spécialistes abandonnent leurs fonctions de contrôle. Ces brefs épisodes d'interruption et d'augmentation de vigilance ne sont pas en eux-mêmes des épisodes « d'attention consciente » comme chez les humains (comme on le dit de façon redondante) ou tout au moins, ce ne sont pas nécessairement des instances de ce type d'état, mais ce sont probablement les précurseurs nécessaires, dans l'évolution, de nos états conscients.

Neumann spécule que ces réponses d'orientation ont commencé comme des réactions à des signaux d'alarme, mais se sont révélées si utiles pour provoquer une mise à jour généralisée que les animaux ont commencé à se placer en position d'orientation de plus en plus souvent... Leurs systèmes nerveux avaient besoin d'une modalité « Tout le monde sur le pont ! », mais une fois

que celle-ci a été mise en place, cela ne coûtait rien de la mettre en marche plus souvent, et elle a touché ses dividendes convenablement pour fournir des informations meilleures sur les états de l'environnement, ou sur les états propres de l'animal. On peut dire que c'est devenu une habitude, qui n'était plus seulement sous le contrôle de stimuli externes, mais était initiée de façon interne (comme des entraînements aux signaux d'incendie).

La vigilance régulière s'est transformée par étapes en *exploration* régulière, et une nouvelle stratégie comportementale a commencé à évoluer : la stratégie d'acquisition de l'information « pour elle-même », juste au cas où cela se révélerait utile un jour. La plupart des mammifères ont été conduits à cette stratégie, en particulier les primates, qui ont développé des yeux très mobiles qui, par saccades, fournissaient un balayage presque ininterrompu du monde alentour. Cette étape a marqué un déplacement assez fondamental dans l'économie des organismes qui l'ont franchie : la curiosité, l'appétit épistémique est apparu. Plutôt que de recueillir de l'information seulement sur une base payez-tout-de-suite-et-consommez-immédiatement, ils ont commencé à devenir ce que le psychologue George Miller a appelé des *informativores* : des organismes qui ont faim d'autres informations sur le monde qu'ils habitent (et sur eux-mêmes). Mais ils n'ont pas inventé ni déployé de systèmes entièrement nouveaux de saisie de l'information. Comme c'est d'habitude le cas dans l'évolution, ils ont assemblé ces nouveaux systèmes à partir de l'équipement que leur héritage leur avait déjà fourni. Cette histoire a laissé des traces, en particulier sur les tonalités émotionnelles et affectives de la conscience, car même si les animaux supérieurs ont fini par devenir des collecteurs « désintéressés » d'information, leurs « rapporteurs » étaient seulement les signaux d'alarme et les meneurs de ban de leurs ancêtres, n'envoyant jamais un message « directement », mais déposant toujours un « tissage » éditorial positif ou négatif sur l'information qu'ils fournissaient. Enlevons les guillemets et les métaphores : les liens innés des états informationnels à l'abandon et à l'acceptation, à l'évitement et aux renforcements n'ont pas été rompus, mais seulement atténués et redirigés. (Nous reviendrons sur ce point au chapitre 11.)

Chez les mammifères, ce développement évolutif a été promu par une division du travail dans le cerveau, qui a créé deux aires spécialisées : en gros, l'aire *dorsale* et l'aire *ventrale*. (Ce qui suit

reprend l'hypothèse du neuropsychologue Michael Kinsbourne.) Le cerveau dorsal s'est vu accorder des responsabilités de *pilotage* « direct » pour garder le vaisseau (le corps de l'organisme) à l'abri des dangers ; comme les routines « de détection de collision » qu'on trouve dans les jeux vidéo, il devait continuellement balayer le champ des objets approchant et s'en allant ; il était en général chargé d'éviter à l'organisme de se cogner dans des objets ou de tomber du haut des falaises. Le cerveau ventral a donc disposé d'un peu de temps libre pour se concentrer sur l'identification des divers objets du monde ; il a pu se permettre de se focaliser sur les objets particuliers et de les analyser de façon relativement lente, sériee, parce qu'il pouvait compter sur le système dorsal pour garder le vaisseau à l'abri des écueils. Chez les primates, comme le spécule Kinsbourne, cette spécialisation dorsale-ventrale s'est altérée ; elle a évolué plus avant sous la forme des spécialisations célèbres de l'hémisphère gauche et de l'hémisphère droit : l'hémisphère droit est global et spatio-temporel, et l'hémisphère gauche plus concentré, analytique et sériel.

Nous n'avons exploré qu'un seul quartier dans l'histoire évolutive des systèmes nerveux, et nous nous sommes concentrés sur le mécanisme évolutionnaire le plus fondamental : la sélection de *génotypes* particuliers (combinaisons de gènes) qui se sont révélés produire des individus mieux adaptés (*phénotypes*) que les autres génotypes possibles. Les organismes qui ont eu la chance d'être mieux câblés à la naissance tendent à produire plus de progéniture capable de survivre ; c'est ainsi que le bon câblage se répand à travers la population. Et nous avons esquissé une progression de l'espace de conception des détecteurs de choses bonnes ou mauvaises les plus élémentaires à des collections de mécanismes de ce type organisés en une architecture dotée d'une capacité considérable de production d'anticipations utiles dans des environnements relativement stables et prévisibles.

Pour la phase suivante de notre histoire, nous devons introduire une innovation majeure : l'émergence de phénotypes individuels dont l'intérieur n'est pas entièrement câblé de façon innée, mais plutôt variables ou *plastiques*, et qui par conséquent peuvent apprendre, durant leur vie. L'émergence de la plasticité dans les systèmes nerveux est survenue (en gros) au même moment que les développements que nous avons déjà esquissés, et elle a produit deux nouveaux véhicules dans lesquels l'évolution a pu prendre place, à une vitesse bien plus grande que l'évolution

génétique qui est sans assistance parce qu'elle passe par la mutation des gènes et la sélection naturelle. Étant donné que certaines des caractéristiques complexes de la conscience humaine sont le résultat des développements qui se sont produits, et qui continuent à se produire dans ces nouveaux véhicules, il nous faut avoir une idée claire, bien qu'élémentaire, de leurs relations qu'ils entretiennent entre eux et vis-à-vis du processus sous-jacent de l'évolution génétique.

L'évolution dans le cerveau et l'effet Baldwin

Nous partons tous du principe que le futur ressemblera au passé. Bien qu'elle soit indémontrable, c'est la prémisse essentielle sur laquelle s'appuie toutes nos inférences inductives, comme l'a fait remarquer Hume. Dame Nature (l'agent intégré au processus de sélection naturelle qui développe les conceptions) part du même principe. À bien des égards, les choses restent les mêmes : la gravitation continue à exercer sa force, l'eau continue à s'évaporer, les organismes continuent à avoir besoin de se remplir à nouveau et à protéger leur corps fait d'eau, les choses qui passent continuent à sous-tendre des portions de plus en plus grandes de leurs rétines, et ainsi de suite. Quand on a affaire à des généralités de ce genre, Dame Nature fournit des solutions à long terme aux problèmes : des détecteurs innés, à base de gravitation, de ce qui est au-dessus et en dessous, des signaux d'alarme innés pour la soif, des circuits innés pour-plonger-quand-quelque-chose-approche. D'autres choses changent, mais de façon prévisible, selon des cycles, et Dame Nature leur répond avec d'autres dispositifs innés : le pelage de certains animaux augmente pendant l'hiver sous l'effet des changements de température, des réveille-matin internes gouvernent les cycles de veille et de sommeil des animaux diurnes et nocturnes. Mais il arrive que Dame Nature – ou qui que ce soit d'autre – ne puisse pas totalement prédire les circonstances et les vicissitudes de l'environnement – quand ceux-ci sont influencés par des processus de nature *chaotique* (Dennett, 1984a, p. 109 *sq.*). Dans ce cas, aucune conception stéréotypée ne pourra s'adapter à toutes les éventualités, si bien que les organismes les plus adaptés seront ceux qui peuvent *se réorganiser* eux-mêmes à un certain degré pour répondre aux

circonstances qu'ils rencontrent. Cette réorganisation est appelée tantôt *apprentissage* tantôt *développement*. La ligne de partage entre ces deux notions est problématique. Est-ce que les oiseaux *apprennent* à voler ? Apprennent-ils à chanter leurs chansons ? (Nottebohm, 1984 ; Marler et Sharman, 1983). À faire pousser leurs ailes ? Est-ce que les bébés *apprennent* à marcher ou à parler ? Étant donné que la ligne de partage (s'il y en a une) est non pertinente pour ce qui nous concerne ici, appelons *fixation postnatale d'organisation* (*postnatal design fixing*) tout processus qui va ainsi du fait d'apprendre à focaliser son regard à apprendre la mécanique quantique. À partir du moment où on naît, il reste de la place pour des variations et celles-ci se trouvent finalement fixées par un processus ou par un autre en un élément relativement permanent de conception pour le reste de la vie. (À partir du moment où vous avez appris à faire du vélo, ou à parler russe, il tend à rester en vous.)

Comment un tel processus de fixation postnatale d'organisation peut-il se réaliser ? D'une seule façon (rien de miraculeux là-dedans) : par un processus fortement analogue aux processus qui fixe les organisations prénatales, ou, en d'autres termes, un processus d'évolution par sélection naturelle qui se produit à l'intérieur de l'individu (dans le phénotype). Quelque chose de déjà fixé au sein de l'individu par la sélection naturelle doit jouer le rôle de sélecteur mécanique, et d'autres choses doivent jouer le rôle de la multitude des candidats à la sélection. On a proposé de nombreuses théories différentes de ce processus, mais toutes – mises à part celles qui sont simplement bizarres ou franchement mystérieuses – ont cette structure et ne diffèrent que sur les détails des mécanismes. Pendant des années, au XX^e siècle, la théorie la plus influente fut le behaviorisme de B.F. Skinner, dans laquelle les couplages de stimuli et de réponses étaient les candidats à la sélection naturelle ; les stimuli « de renforcement » étaient les mécanismes de sélection. Le rôle des stimuli de plaisir ou de douleur dans la formation du comportement – la carotte et le bâton – est indéniable, mais on a largement admis que le mécanisme behavioriste de « conditionnement opératoire » était trop simple pour expliquer les complexités de la fixation postnatale d'organisation chez des espèces aussi compliquées que les êtres humains (et probablement chez les pigeons aussi, mais c'est une autre affaire). Aujourd'hui, on se concentre sur diverses théories qui transportent le processus d'évolution au sein du

cerveau (Dennett, 1974). Différentes versions de cette idée ont plus ou moins circulé depuis des dizaines d'années ; à présent, la possibilité de tester les modèles rivaux par d'énormes simulations sur ordinateur a permis un débat considérable, qu'il serait sage de ne pas aborder ici ⁶.

Pour ce qui nous occupe, disons seulement que d'une manière ou d'une autre, le cerveau plastique est capable de se réorganiser lui-même en s'adaptant en réponse aux éléments nouveaux que l'organisme rencontre dans son environnement ; le processus par lequel le cerveau y parvient doit être un processus mécanique tout à fait analogue à celui de la sélection naturelle. C'est là le premier véhicule nouveau de l'évolution : la fixation postnatale d'organisation dans les cerveaux individuels. Les candidats à la sélection sont diverses structures cérébrales qui contrôlent ou influencent les comportements, et la sélection s'accomplit à travers un processus mécanique d'élimination qui est lui-même installé génétiquement dans le système nerveux.

Chose étonnante, cette capacité, qui est elle-même un produit de l'évolution génétique par sélection naturelle, ne donne pas seulement aux organismes qui la possèdent une longueur d'avance sur leurs cousins au câblage inné qui ne peuvent pas se reconcevoir eux-mêmes, mais se répercute sur le processus de l'évolution génétique et *l'accélère*. C'est un phénomène qui est connu sous différents noms, en particulier celui d'*effet Baldwin* (voir Richards, 1987 ; Schull, 1990). Voici comment il fonctionne.

Considérez une population d'une espèce particulière dans laquelle il y a une variation considérable à la naissance quant à la manière dont les cerveaux sont câblés. Supposons que l'un des câblages possibles donne à son possesseur un Bon Truc – un talent comportemental qui le protège ou qui favorise ses chances de manière radicale. Nous pouvons représenter ce phénomène dans ce que l'on appelle un paysage adaptatif ; la hauteur représente la *fitness* (plus c'est haut, mieux c'est) ; la longitude et la latitude représentent des variables dans le câblage (nous n'avons pas besoin de les spécifier pour cette expérience de pensée).

Comme cela se voit clairement sur la figure, un seul câblage se trouve favorisé ; les autres, aussi « proches » puissent-ils être du bon câblage, sont à peu près à égalité en *fitness*. Cette aiguille dans la botte de foin peut être pratiquement invisible à la sélection naturelle. Même si quelques individus chanceux sont câblés de cette façon, les possibilités que leur fortune se répande sur la

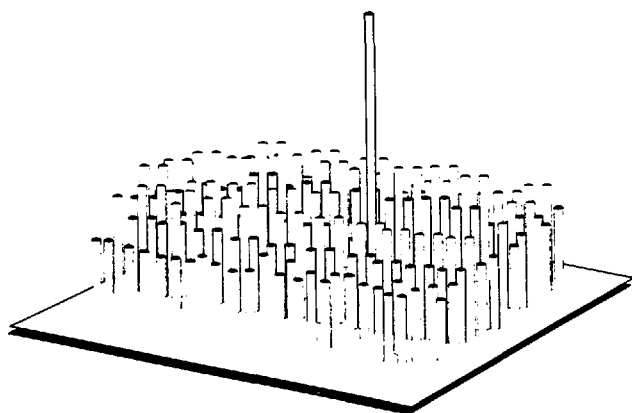


Figure 6.1

population des générations suivantes peuvent être étonnamment faibles à moins qu'il n'y ait une plasticité de conception parmi les individus.

Supposez, par conséquent, que les individus débutent tous en étant différents génétiquement, mais qu'au cours de leur existence, ils se promènent dans l'espace des différentes possibilités d'organisation qui leur sont autorisées, en raison de leur plasticité. Et en raison des circonstances particulières de l'environnement, ils tendent tous à graviter en direction d'un câblage favorisé. Il existe un Bon Truc à apprendre dans leur environnement, et il tendent tous à l'apprendre. Supposez que ce soit un truc tellement bon que ceux qui ne l'apprennent pas soient sérieusement désavantagés, et supposez que ceux qui tendent à ne jamais l'apprendre soient ceux qui débutent leur existence avec des organisations qui sont bien plus éloignées dans l'espace des organisations possibles (et qui par conséquent ont besoin d'une réorganisation postnatale) que ceux qui sont près du Bon Truc.

Une fiction (empruntée à Hinton et Nowlan, 1987) nous aidera à imaginer tout cela. Supposez qu'il y ait dix lieux dans le cerveau d'un animal où un « câble » puisse être attaché de deux façons possibles, A ou B. Supposez que le Bon Truc soit l'organisation qui correspond au câblage AAABBBAAAA, et que tous les autres câblages soient tous également aussi peu intéressants, comportementalement parlant. Puisque toutes ces organisations sont

plastiques, chaque animal, durant son existence, peut essayer l'une quelconque des 2^{10} combinaisons différentes des câblages de A et de B. Les animaux qui sont nés dans un état tel que BAABBBAAAA ne sont qu'à un recâblage de distance du Bon Truc (bien qu'ils puissent, bien entendu, essayer au hasard une série d'autres recâblages en premier lieu). D'autres, qui commencent à essayer des câblages comme BBBAABBBB doivent réaliser au moins dix recâblages (sans faire jamais de mauvais recâblages) avant de trouver le Bon Truc. Les animaux dont les cerveaux partent plus près de la cible ont un avantage de survie sur ceux qui partent plus loin, même s'il n'y a pas d'avantage sélectif *supplémentaire* du fait que l'on soit né avec une structure « qui manque la cible de peu » par opposition à une structure « qui manque la cible de loin » (comme on le voit bien avec la figure 6.1). La population de la génération suivante tendra alors à consister principalement en individus plus proches de la cible (et par conséquent plus aptes à trouver la cible durant leurs existences) ; ce processus peut se poursuivre jusqu'à ce que l'ensemble de la population soit *génétiquement* fixée sur le Bon Truc. Un Bon Truc « découvert » par des individus de cette façon peut ainsi être repassé, relativement aisément, aux générations futures.

Si nous donnons aux individus des possibilités variables de tomber sur le Bon Truc (et puis de le « reconnaître » et de « s'y tenir ») durant leur existence, l'aiguille quasi invisible dans la botte de foin de la figure 6.1 devient le sommet d'une crête tout à fait visible que la sélection naturelle peut grimper (figure 6.2). Ce processus, l'effet Baldwin, peut de prime abord paraître ressembler beaucoup à l'idée lamarckienne aujourd'hui discréditée de la transmission génétique des caractères acquis, mais ce n'est pas la même idée. Rien de ce que l'individu apprend n'est transmis à sa progéniture. Il se trouve seulement que les individus qui ont la chance d'être plus proches dans l'espace exploratoire de conception d'un Bon Truc à apprendre auront tendance à avoir plus de progéniture ; ils seront plus près du Bon Truc. À travers les générations, la compétition devient plus dure : finalement, à moins que vous ne soyez né avec le Bon Truc (ou presque avec lui), vous ne serez pas assez près pour entrer dans la compétition. S'il n'était pas destiné à produire la plasticité l'effet ne serait pas là, car « manquer de près ou de loin c'est toujours manquer » à moins que vous n'essayiez de nouvelles variations jusqu'à ce que vous obteniez le bon résultat.

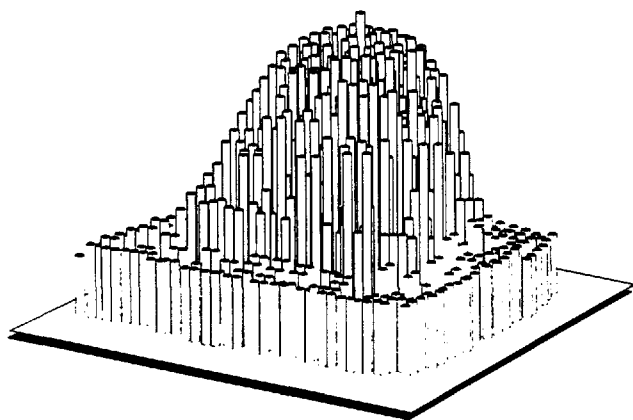


Figure 6.2

Grâce à l'effet Baldwin, on peut dire que les espèces testent d'avance l'efficacité de différentes organisations particulières à travers des explorations phénotypiques individuelles dans l'espace des possibilités proches. Si un dispositif gagnant particulier se trouve ainsi découvert, cette découverte *créera* une nouvelle pression sélective : les organismes qui sont plus proches de cette découverte dans le paysage adaptatif auront un avantage clair sur ceux qui se situent plus loin. Cela veut dire que les espèces douées de plasticité *tendront* à évoluer plus rapidement (et avec plus de « lucidité ») que celles qui n'en ont pas. Ainsi l'évolution dans le second véhicule, la plasticité phénotypique, peut-elle faire progresser l'évolution dans le premier véhicule, celui de la variation génétique. (Nous allons bientôt voir un contre-effet produit par des interactions avec le troisième véhicule.)

La plasticité dans le cerveau humain : le décor

« L'entendement par sa propre force innée se forge des outils pour d'autres œuvres intellectuelles et grâce à ces œuvres, d'autres outils, c'est-à-dire le pouvoir de chercher plus avant. Ainsi avance-t-il par degrés jusqu'au faite de la sagesse ⁷. »

Spinoza, 1677.

Les systèmes nerveux qui sont câblés de façon innée sont de faible poids et sont producteurs d'énergie ; et ils conviennent bien aux organismes qui ont affaire à des environnements stéréotypiques et disposent de budgets limités. Des esprits plus imaginatifs, grâce à leur plasticité, sont capables non seulement d'anticipations stéréotypées, mais aussi de s'ajuster à des courants perçus. Même le modeste crapaud dispose d'un certain degré de liberté dans la manière dont il répond aux nouvelles situations, altérant lentement les structures d'activités qui lui permettent de détecter – avec des délais considérables – ceux des changements de son environnement qui sont les plus importants pour son bien-être (Ewert, 1987). Dans le cerveau du crapaud, un dispositif permettant d'interagir avec le monde évolue à une vitesse qui est bien plus considérable que celle de la sélection naturelle – avec des « générations » qui durent des secondes ou des minutes, pas des années. Mais pour un contrôle vraiment puissant, ce qui est requis, c'est une machine anticipatrice qui s'ajuste essentiellement en quelques millisecondes ; pour cela, il faut un dispositif virtuose producteur de futur, un système qui puisse prévoir, éviter les ornières dans sa propre activité, résoudre les problèmes avant de les rencontrer et reconnaître des signes avant-coureurs entièrement nouveaux de bonnes et de mauvaises choses. En dépit de toute notre stupidité, nous autres êtres humains sommes considérablement mieux équipés pour cette tâche que n'importe quel autre être se contrôlant lui-même et c'est notre énorme cerveau qui rend cela possible. Mais comment ?

Passons en revue les progrès réalisés. Nous avons esquissé une histoire – qui ne constitue qu'un quartier de la fabrique mul-

tidimensionnelle de l'histoire évolutionnaire – de l'évolution du cerveau des primates. Fondée sur des millénaires de systèmes nerveux antérieurs, elle consiste en un conglomerat de circuits spécialisés construits pour accomplir des tâches particulières dans l'économie des ancêtres primates : des détecteurs d'objets qui passent câblés à des mécanismes d'évitement, des détecteurs quel-qu'un-me-regarde câblés à des discriminateurs d'amis-ou-ennemis eux-mêmes câblés à leurs propres sous-routines. Nous pouvons aussi ajouter des circuits spécifiques aux primates tels que les circuits de coordination de la main avec l'œil destinés à attraper des baies ou des graines, d'autres destinés à attraper des branches ou même à entrer en relation avec des objets proches du visage (Rizzolati, Gentiluci, et Matelli, 1985). Grâce à des yeux mobiles et un penchant pour l'exploration et la mise à jour, ces cerveaux de primates ont été régulièrement inondés d'information multimédia (comme diraient les chercheurs en neurosciences). Par conséquent, un nouveau problème s'est posé pour eux : le problème du contrôle de haut niveau.

Un problème posé est aussi une opportunité, une porte qui ouvre une nouvelle portion de l'espace d'organisation. Jusqu'à présent, nous pouvons le supposer, les systèmes nerveux ont résolu le problème « Que faire maintenant ? » par un acte de balancement relativement simple entre un répertoire strictement limité d'actions – sinon les fameuses quatre fonctions (se battre, fuir, se nourrir, ou se trouver un compagnon), du moins une modeste proportion de celles-ci. Mais à présent, que la plasticité fonctionnelle s'est accrue et que nous disposons de plus d'information « centralisée » venue de tous les divers systèmes spécialisés, le problème de savoir quoi faire maintenant est l'origine d'un métaproblème : *à quoi penser ensuite ?* C'est très bien de s'équiper d'une sous-routine « Tout le monde sur le pont ! », mais ensuite, une fois que tout le monde est sur le pont, on doit avoir une façon de se débrouiller avec le flux des volontaires. Nous ne devrions pas nous attendre à avoir un capitaine déjà disponible (qu'aurait-il pu bien faire jusque-là ?), en sorte que les conflits entre volontaires ont dû se régler sans officier supérieur. (Comme nous l'avons déjà vu, dans l'exemple du système immunitaire, des actions concertées, organisées ne doivent pas toujours dépendre du contrôle à partir d'un commandement supérieur.) Le modèle pionnier de ce genre de processus est dû à Oliver Selfridge (1959) : c'est celui de l'architecture du Pandémonium en intelligence

artificielle, dans lequel de nombreux « démons » luttaienent en parallèle pour obtenir l'hégémonie, et puisque le nom donné par Selfridge à cette sorte d'architecture convient si bien, je l'utiliserai dans ce livre en un sens générique, pour faire référence à celle-ci et à ses descendants, directs et indirects, tels que « le programme de conflits » (Norman and Shallice, 1980, Shallice, 1988) et les réseaux d'IA Qui Gagne Rafle Tout (*Winner Take All*) de Ballard et Feldman (1982) et ses descendants.

Le programme de conflits dans le style Pandémonium, mené directement par les traits présents de l'environnement, a néanmoins besoin d'un système nerveux doué d'une capacité limitée de projection vers le futur. Odmar Neuman a formé l'hypothèse que des réactions d'orientation, à l'origine motivées par des nouveautés intervenues dans l'environnement, vinrent à être produite de façon endogène (de l'intérieur) ; de même, nous pouvons former l'hypothèse qu'il existait une pression en faveur du développement d'une façon plus endogène de résoudre le métaproblème de savoir quoi penser ensuite, une pression en vue de créer quelque chose sur l'intérieur qui aurait un peu plus les pouvoirs organisationnels qu'on imagine être ceux d'un capitaine.

Voyons à quoi pouvait bien ressembler à ce moment le comportement de notre ancêtre primate hypothétique vu de l'extérieur (nous différons l'examen des questions portant sur l'effet que cela ferait d'être un tel primate) : un animal capable d'apprendre de nouveaux trucs, presque sans cesse vigilant et sensible aux nouveautés, et doté d'un « espace d'attention à court terme » et d'une tendance à voir son attention « retenue » par des traits de l'environnement susceptibles de le distraire. Pas de projets à long terme pour cet animal, tout au moins pas de projets nouveaux. (Il nous faut laisser une place pour des sous-routines stéréotypées de longue durée câblées génétiquement, telles que les routines de fabrication de nids des oiseaux, de construction de barrages pour les castors, ou de recherche de nourriture chez les oiseaux et les écureuils.)

À partir de ces systèmes nerveux de base, nous voulons à présent imaginer l'organisation d'un esprit plus humain, doué de quelque chose comme un « flux de conscience » capable de produire les types de « suites de pensées » sophistiqués dont semble dépendre la civilisation humaine. Les chimpanzés sont l'espèce la plus proche de nous — ils sont génétiquement plus proches, en fait, que les chimpanzés eux-mêmes par rapport aux gorilles

et aux orangs-outans – et on pense couramment que nous avons partagé un ancêtre commun avec les chimpanzés, il y a six millions d'années. Depuis cette rupture décisive, nos cerveaux ont divergé dramatiquement, mais avant tout en taille, plutôt qu'en structure. Alors que les chimpanzés ont un cerveau qui a en gros la même taille que celui de notre ancêtre commun (et il est important, et difficile, de garder à l'esprit que les chimpanzés ont eux aussi évolué à partir de notre ancêtre commun), le cerveau de nos ancêtres hominiens a crû d'au moins quatre fois. Cette augmentation de volume ne s'est pas produite immédiatement, parce que cinq millions d'années après la rupture avec les proto-chimpanzés, nos ancêtres hominiens ont continué à avoir des cerveaux de la taille de ceux des singes, en dépit du fait qu'ils sont devenus bipèdes il y a au moins trois millions et demi d'années. Puis au début de la période de glaciation il y a environ deux millions et demi d'années, la Grande Encéphalisation a commencé, et s'est terminée environ il y a cent cinquante mille ans – *avant* le développement du langage, de la cuisine et de l'agriculture. Pourquoi le cerveau de nos ancêtres a-t-il si vite augmenté en taille (dans le temps propre à l'évolution, c'était plus une explosion qu'une efflorescence) ? On débat beaucoup sur ce point (pour des explications très claires, voir les livres de William Calvin). Mais personne ne conteste la nature du produit : le cerveau de l'*Homo sapiens* à ses débuts (il vivait en gros il y a cent cinquante mille ans à la fin de la plus récente période de glaciation, il n'y a que dix mille ans) était un cerveau extrêmement complexe d'une plasticité sans égale, presque indiscernable du nôtre quant à sa taille et à sa forme. C'est important : la croissance étonnante du cerveau hominien était essentiellement parvenue à son terme *avant* le développement du langage, et en ce sens elle ne peut permettre de répondre aux complexités de l'esprit que le langage a rendu possible. Les spécialisations innées du langage, dont, entre autres, le linguiste Noam Chomsky forme l'hypothèse, commencent à être confirmées par les toutes dernières avancées de la neuroanatomie, exploitent sans doute les circuits séquentiels antérieurs (Calvin, 1989a) et sont accélérées par l'effet Baldwin. De plus, l'expansion la plus remarquable des pouvoirs *mentaux* humains (comme l'atteste le développement de la cuisine, de l'agriculture, de l'art, et, en un mot, de la civilisation) s'est produite tout entière encore plus récemment, depuis la fin de la dernière période de glaciation, en un

clin d'œil de dix mille ans qui est aussi bon qu'instantané du point de vue évolutif, lequel mesure les courants en millions d'années. Notre cerveau est équipé à la naissance d'une petite quantité de pouvoirs qui faisaient totalement défaut à celui de nos ancêtres il y a dix mille ans. Il s'ensuit que l'avance extraordinaire prise par l'*Homo sapiens* durant les dix mille dernières années doit être dû presque tout entier à un modelage de la plasticité de ce cerveau par des voies radicalement différentes et nouvelles – il a créé une sorte de *logiciel* pour augmenter ses pouvoirs sous-jacents (Dennett, 1986).

En bref, nos ancêtres doivent avoir appris certains Bons Trucs qu'ils pouvaient effectuer avec le matériel (*hardware*) ajustable, que notre espèce vient juste de commencer à intégrer dans son génome par l'intermédiaire de l'effet Baldwin. De plus, comme nous le verrons, il est justifié de croire qu'en dépit de la pression sélective initiale en faveur d'un « câblage inné » de ces Bons Trucs, les trucs ont tellement changé la nature de l'environnement pour notre espèce qu'aucune pression sélective significative n'appelle plus d'autres câblages innés. Il est probable que presque toute la pression sélective pesant sur l'organisation du système nerveux humain a été épongée par les effets secondaires de cette nouvelle chance d'organisation qu'ont exploitée nos ancêtres.

Jusqu'ici, j'ai essayé d'éviter de parler des systèmes nerveux plus simples comme *représentant* quoi que ce soit dans le monde. Les diverses conceptions que nous avons considérées, aussi bien plastiques que câblées de façon innée, semblent bien être sensibles à, ou répondre à, ou construites-avec-un œil-sur, ou utiliser, de l'information sur divers traits de l'environnement des organismes ; par conséquent, en ce sens minimal, elles peuvent être appelées des représentations. Mais à présent il nous faut faire une pause pour examiner quels sont les traits de ces organisations complexes qui peuvent nous conduire à les considérer comme des systèmes de représentations.

Une partie de la variabilité du cerveau sert simplement de véhicule aux structures transitoires d'activité cérébrale qui *enregistrent* en quelque sorte ou qui du moins *suivent à la trace* les traits variables qui comptent dans l'environnement. Quelque chose dans le cerveau a besoin de changer s'il peut détecter l'oiseau qui vole aux alentours, ou la baisse de la température ambiante, ou l'un des états de l'organisme lui-même – la baisse en sucre dans le sang, l'augmentation de dioxyde de carbone dans les

poumons. De plus, et c'est le levier qui donne à la véritable représentation son point d'appui – ces structures internes changeantes continuent à être capables de « suivre à la trace » (en un sens large) les traits auxquels ils font référence quand ils sont temporellement coupés du commerce causal avec leurs référents. « Un zèbre qui vient de voir un lion n'oublie pas où se trouve le lion quand il cesse de regarder le lion pendant un instant. Le lion n'oublie pas où se trouve le zèbre » (Margolis, 1987, p. 53). On peut comparer ce phénomène à celui du tournesol qui suit le passage du soleil à travers le ciel, ajustant son angle comme un panneau solaire mobile pour maximiser les rayons de soleil qui tombent sur lui. Si le soleil se trouve temporairement obscurci, le tournesol ne peut pas projeter la trajectoire : le mécanisme qui est sensible au passage du soleil ne représente pas le passage du soleil au sens large. On peut trouver les rudiments de la représentation réelle chez de nombreux animaux inférieurs (et nous ne devrions pas interdire, *a priori*, la possibilité de la représentation réelle dans les plantes), mais chez les humains la capacité à représenter est montée en flèche.

Parmi les choses qu'un cerveau humain adulte peut dans une certaine mesure représenter, on n'a pas seulement :

1. la position du corps et de ses membres ;
2. une tache de lumière rouge ;
3. un certain degré de faim ;
4. un certain degré de soif ;
5. l'odeur d'un excellent vieux bourgogne ;

on a aussi :

6. l'odeur d'un excellent vieux bourgogne *en tant qu'*odeur d'un Chambertin 1971 ;
7. Paris ;
8. Atlantis ;
9. la racine carrée du plus large nombre premier plus petit que 20 ;
10. le concept d'un tire-bouchon nickelé combiné à une désagrafeuse.

Il semble presque certain qu'aucun autre cerveau animal ne peut représenter (6)-(10). Qui plus est, il faut un processus considérable d'ajustement du cerveau de l'enfant humain avant que ces phénomènes puissent être enregistrés ou représentés. Les cinq premiers, par contre, pourraient bien être des choses que presque

n'importe quel cerveau peut représenter (en un sens quelconque) sans entraînement préalable.

D'une manière ou d'une autre, en tout cas, la façon dont le cerveau représente la faim doit différer physiquement de celle dont il représente la soif – puisqu'elle doit gouverner un comportement distinct selon la nature de ce qui est représenté. Il doit aussi, à l'autre extrême de la chaîne, exister une différence entre la manière dont un cerveau humain adulte particulier représente Paris et Atlantis, car penser à l'un n'est pas la même chose que penser à l'autre. Comment un état particulier ou événement dans le cerveau peut-il représenter un trait du monde plutôt qu'un autre⁸ ? Et quel que soit ce qui fait qu'un trait du cerveau représente ce qu'il représente, *comment vient-il* à représenter ce qu'il représente ? Ici encore (j'ai bien peur que ce refrain ne devienne lassant !), il existe toute une gamme de possibilités, mises en place par des processus évolutifs : certains éléments du système de représentation peuvent – et en fait doivent être (Dennett, 1969) – fixés de façon innée, et le reste doit être « appris ». Alors que certaines des catégories qui comptent dans la vie (comme la faim et la soif) nous sont sans aucun doute « données » par un dispositif inné à la naissance, nous devons en développer d'autres nous-mêmes⁹.

Comment nous y prenons-nous ? Probablement par un processus de génération et de sélection de structures d'activité neuronale dans le cortex cérébral, l'énorme manteau plein de circonvolutions qui a rapidement bourgeonné dans le crâne humain et qui couvre à présent complètement l'ancien cerveau animal qu'il recouvre. Dire seulement que c'est un processus évolutif se produisant principalement dans le cortex laisse trop de choses dans l'ombre, et à ce niveau de complexité et de sophistication, même si nous réussissons à expliquer le processus au niveau des synapses ou des groupes de neurones, nous pourrions nous tromper sur d'autres aspects de ce qui se produit. Si nous pouvons comprendre le phénomène, nous devons d'abord passer à un niveau plus général et plus abstrait. À partir du moment où nous comprenons en gros comment se passent les choses au niveau le plus élevé, nous pouvons envisager de descendre à nouveau au niveau le plus mécanique du cerveau.

La plasticité rend l'apprentissage possible, mais il vaudrait mieux qu'il y ait *quelque chose à apprendre* dans l'environnement extérieur qui soit déjà le produit d'un processus de conception

préalable, de façon à ce que chacun de nous n'ait pas à réinventer la roue. L'évolution culturelle, et la transmission de ses produits, sont le second nouveau véhicule de l'évolution ; elle dépend de la plasticité phénotypique d'une façon très comparable à celle dont la plasticité phénotypique dépend de la variation génétique. nous autres êtres humains avons utilisé notre plasticité non seulement pour apprendre, mais pour apprendre comment apprendre mieux, pour ensuite apprendre mieux comment apprendre mieux, et ainsi de suite. Nous avons aussi appris comment mettre les fruits de cet apprentissage à la disposition des novices. Nous *installons*, pour ainsi dire, un *système* d'habitudes déjà inventé et dans une large mesure « débarrassé de bogues » dans le cerveau en partie non structuré.

L'invention de bonnes et de mauvaises habitudes d'autostimulation

« Comment puis-je dire ce que je pense tant que je ne vois pas ce que je dis ? »

E.M. Forster, 1960.

« Nous parlons, non pas seulement pour dire aux autres ce que nous pensons, mais pour nous dire à nous-mêmes ce que nous pensons. »

J. Hughlins Jackson, 1915.

Comment ce partage du logiciel a-t-il pu survenir ? On peut répondre « tout simplement comme ça ». Envisageons un moment dans l'histoire des premiers *Homo sapiens* où le langage – ou plutôt ce que nous devrions appeler un proto-langage – commençait seulement à se développer. Ces ancêtres étaient des omnivores bipèdes, vivant dans de petits groupes de parenté étroite, qui avaient probablement développé des habitudes de vocalisation pour des objectifs particuliers comparables à celles des chimpanzés et des gorilles, et même d'espèces plus distantes encore, telles que les singes vervets (Cheney and Seyfarth, 1990). Nous pouvons supposer que les actes de communication (ou de quasi-communication) exécutés par ces vocalisations n'étaient pas encore des actes de langage à part entière (Bennett, 1976) dans lesquels l'intention du Locuteur de produire un certain effet sur l'Au-

ditore dépend de la façon dont l'Auditoire évalue cette intention même ¹⁰. Mais nous pouvons supposer que ces ancêtres, comme les primates contemporains qui recourent à des vocalisations, distinguaient différents énonciateurs et différents auditeurs dans différentes occasions, en utilisant des informations sur ce que les deux parties pourraient croire ou vouloir ¹¹. Par exemple, l'hominien Alf ne s'inquiéterait pas d'essayer de faire croire à l'hominien Bob qu'il n'y a pas de nourriture dans la caverne (en grognant « Padbouffici ») si Alf croyait que Bob savait déjà qu'il n'y avait pas de nourriture dans la caverne. Et si Bob pensait qu'Alf voulait l'induire en erreur, Bob serait prêt à considérer les vocalisations d'Alf avec prudence et scepticisme ¹².

Or on peut spéculer qu'il est souvent arrivé que quand l'un de ces hominiens était aux prises avec un problème qu'il ne pouvait pas résoudre, il « demandait de l'aide », et en particulier il « demandait de l'information ». Quelquefois, l'Auditoire présent répondait en « communiquant » à l'enquêteur quelque chose qui avait l'effet voulu, le faisant sortir de son excitation, ou lui faisant « voir » la solution de son problème. Pour que cette pratique puisse se répandre dans la communauté, ceux qui demandaient de l'aide devaient être capables de répondre à l'occasion en jouant le rôle de donneurs de réponse. Ils devaient avoir la capacité comportementale d'être provoqués à faire des énonciations occasionnellement « utiles » quand ils étaient soumis à des énonciations de « requête » de la part des autres. Par exemple, si un hominien savait quelque chose et qu'on lui « posait une question » à ce sujet, cela pouvait avoir l'effet normal – bien que sujet à des exceptions – de le conduire à « dire ce qu'il savait ».

En d'autres termes, je propose l'idée qu'il y a eu un temps, au cours de l'évolution du langage, où les vocalisations avaient pour fonction de susciter et de partager des informations utiles, bien qu'on ne puisse supposer qu'un esprit coopératif d'aide mutuelle aurait une valeur de survie ou serait un système stable s'il émergeait. (Voir par exemple Dawkins, 1982, p. 55 *sq.* ; voir aussi Sperber et Wilson, 1986.) Nous devons plutôt supposer que les coûts et les bénéfices résultant de la participation à une telle pratique étaient en quelque sorte « visibles » pour ces créatures ; un nombre suffisant d'entre elles voyaient les bénéfices qu'elles pourraient *elles-mêmes* en tirer comme compensant les coûts en sorte que les habitudes de communication deviennent établies dans la communauté.

Puis, un beau jour (selon cette reconstruction rationnelle), l'un de ces hominiens demanda de l'aide « par erreur » à un moment où il n'y avait pas d'auditoire secourable à portée d'oreille – sauf lui-même ! Quand il entendit sa propre requête, la stimulation provoqua juste la sorte d'énonciation susceptible d'aider les autres que celle qu'aurait causé une requête venue d'un autre. Et à son grand plaisir, la créature en question découvrit qu'elle s'était provoquée elle-même à répondre à sa propre question.

Par cette expérience de pensée délibérément simpliste, j'essaie de justifier l'idée que la pratique consistant à se poser à soi-même des questions a pu surgir comme un effet naturel secondaire de l'activité de poser des questions aux autres, et que son utilité serait similaire : ce serait un comportement qui pourrait être reconnu comme améliorant la situation d'un individu en promouvant une action guidée plus efficace. Pour que cette pratique ait cette utilité, il suffit que les relations d'accès préexistantes à *l'intérieur* du cerveau d'un individu soient moins qu'optimales. Supposez, en d'autres termes, que bien que l'information correcte acquise pour un objectif particulier soit déjà *dans le cerveau*, elle ne soit pas aux mains des bons spécialistes ; le sous-système dans le cerveau qui a besoin de l'information ne peut pas l'obtenir directement à partir du spécialiste – parce que l'évolution n'a pas réussi à produire un tel « câble ». Mais provoquer le spécialiste à « diffuser » l'information dans l'environnement, puis s'appuyer sur une paire existante d'oreilles (et un système auditif) serait une manière de construire un « câble virtuel » entre les sous-systèmes pertinents ¹³.

Un tel acte d'autostimulation pourrait frayer un nouveau chemin important entre les composants internes d'un individu. En gros, pousser une information à travers ses oreilles et son système auditif peut bien stimuler précisément les sortes de connexions que l'on recherche, peut déclencher précisément les mécanismes associatifs corrects, provoquer juste le bon morceau de choix mental au bout de la langue. On peut alors le dire, et ainsi obtenir la réponse que l'on espérait.

À partir du moment où des habitudes grossières d'autostimulation commencèrent à s'établir comme de Bons Trucs dans le comportement des populations hominiennes, nous pourrions nous attendre à ce qu'elles se soient rapidement raffinées, à la fois dans les habitudes comportementales apprises de la population et, grâce à l'effet Baldwin, dans les prédispositions géné-

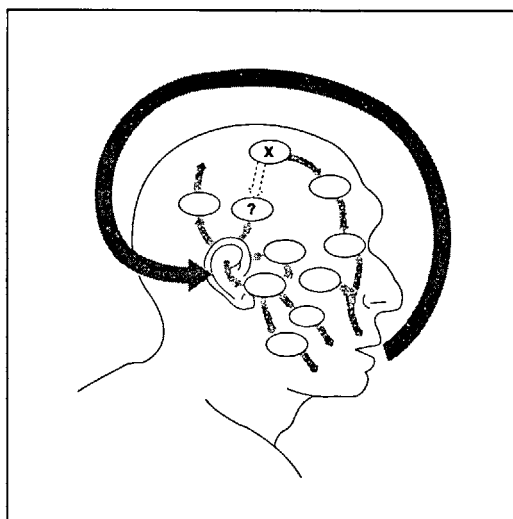


Figure 6.3

tiques et les améliorations ultérieures d'effcience et d'efficacité. En particulier, nous pouvons spéculer que les plus grandes vertus du discours *sotto voce* à soi-même seraient reconnues, conduisant plus tard à un discours entièrement silencieux adressé à soi-même. Le processus silencieux maintiendrait la boucle de l'autostimulation, mais renoncerait à la vocalisation périphérique et aux portions d'audition du processus, qui ne contribuait pas pour grand-chose. Cette innovation aurait un autre avantage, accepté avec opportunisme, de permettre un certain caractère privé pour la pratique de l'autostimulation cognitive. (Dans le prochain chapitre, nous considérerons la manière dont ces lignes de communication écourtées pourraient fonctionner.) Ce caractère privé serait particulièrement utile au moment où des membres de la même espèce capables de comprendre se trouveraient à portée d'oreille. Ce comportement privé de dialogue avec soi-même pourrait bien ne pas être la meilleure façon possible de modifier l'architecture fonctionnelle existante du cerveau, mais cela sera une amélioration qu'on a directement sous la main, aisément découverte, et cela pourrait plus que suffire. Elle serait lente et laborieuse, comparée aux processus cognitifs inconscients rapides sur lesquels il était basé, parce qu'il devait faire usage de grandes portions du système nerveux « conçus pour d'autres

objectifs » – en particulier pour la production et la compréhension du discours audible. Il serait tout aussi linéaire (limité à un sujet à la fois) que la communication sociale à partir de laquelle il a évolué. Et il serait dépendant, tout au moins au départ, des catégories informationnelles incorporées dans les actions qu'il exploite. (Si l'hominien ne pouvait « dire » que cinquante choses à ses congénères, il pourrait y avoir cinquante choses qu'il pourrait se dire à lui-même.)

Parler tout haut n'est qu'une possibilité parmi d'autres. Se dessiner des images à soi-même est un autre acte d'automanipulation qui se laisse apprécier immédiatement. Supposez qu'un jour l'un de ces hominiens ait dessiné pour s'amuser deux lignes parallèles sur le sol de sa caverne et qu'à la vue du résultat, ces deux lignes lui aient rappelé visuellement les rives parallèles de la rivière qu'il aurait eu à traverser plus tard dans la journée. Supposons que cela lui ait rappelé qu'il lui fallait emmener sa liane pour la traverser. S'il n'avait pas dessiné « l'image », nous pouvons supposer qu'il aurait marché jusqu'à la rivière, *puis* compris, après un rapide coup d'œil, qu'il avait besoin de sa corde ; il aurait ainsi dû revenir sur ses pas. Cela pourrait constituer une économie de temps et d'énergie qui pourrait provoquer de nouvelles habitudes, puis se développer en dernière instance sous la forme d'une activité *privée* consistant à dessiner des diagrammes « dans son esprit ».

Le don qu'ont les humains d'inventer de nouvelles voies de communication internes se révèle à l'occasion très nettement dans les cas de lésions cérébrales. Les gens ont des capacités extraordinaires à surmonter les lésions cérébrales, et ce n'est *jamais* une affaire de « cicatrisation » ou de réparation des circuits endommagés. Ils découvrent plutôt de nouvelles manières de réussir les anciens trucs et l'exploration active joue un grand rôle dans la réhabilitation. Les recherches sur les patients au cerveau divisé (Gazzaniga, 1978) fournissent une anecdote particulièrement suggestive à cet égard. Les hémisphères gauche et droit sont habituellement connectés par un large pont de fibres que l'on appelle le corps calleux. Quand on coupe celui-ci chirurgicalement (dans le traitement des épilepsies graves), les deux hémisphères perdent leurs principales « lignes » de connexion directe et se retrouvent pratiquement tenus au secret. Si on demande à l'un de ces patients d'identifier un objet – par exemple un stylo – en essayant de le chercher à tâtons dans un sac, sa

réussite dépendra de la main qu'il utilise pour cette tâche. La majeure partie du câblage du corps est disposée *de façon contralatérale*, l'hémisphère gauche obtenant son information – et contrôlant – la partie gauche du corps, et vice versa. Puisque l'hémisphère gauche contrôle normalement le langage, quand le patient plonge sa main droite dans le sac, il peut aisément dire ce qu'il y a dans le sac ; mais s'il fait la même chose avec sa main gauche, seul l'hémisphère droit obtient l'information que l'objet est un stylo ; ce patient est donc incapable de diriger la voix pour exprimer ce fait. Mais il semble parfois, que l'hémisphère droit découvre un habile stratagème : en trouvant le bout du stylo et en le piquant dans sa paume, il produit un signal de douleur aigu qui est envoyé au bras gauche et certaines fibres transmettrices de douleur se trouvent être câblées *ipselatéralement*. L'hémisphère gauche, celui qui contrôle le langage, obtient ainsi un indice : c'est quelque chose d'assez pointu pour causer une douleur. « C'est pointu – c'est peut-être un stylo ou un crayon ? » L'hémisphère droit, entendant cette vocalisation, peut l'aider avec quelques indices, froncer les sourcils pour *stylo*, sourire pour *crayon* – en sorte qu'au bout d'une brève série de « Vingt Questions », l'hémisphère droit est conduit à la réponse correcte. On rapporte quantité d'anecdotes au sujet de ce genre de tests ingénieux inventés sur le moment par des patients au cerveau divisé, mais nous devons les envisager avec prudence. Elles peuvent être ce qu'elles semblent être : des cas qui montrent l'habileté avec laquelle le cerveau peut découvrir et implanter des stratégies autostimulatoires pour améliorer ses communications internes en l'absence des câblages « désirés ». Mais elles peuvent être aussi des fictions enjolivées inventées sans s'en rendre compte par des chercheurs qui espèrent obtenir ce genre de données. C'est tout le problème avec les anecdotes.

Nous pourrions nous amuser à rêver à d'autres scénarios plausibles expliquant l'« invention » de modes d'autostimulation utiles, mais cela risquerait d'obscurcir l'idée qui veut que toutes les inventions de ce genre soient utiles à la survie de l'individu. À partir du moment où une habitude générale d'autostimulation exploratoire a été ainsi inculquée, elle pourrait bien engendrer une quantité de variations non fonctionnelles (sans qu'elles soient particulièrement dysfonctionnelles). Il existe, après tout, de nombreuses variétés d'autostimulation et d'automanipulation qui ne semblent apparemment pas avoir d'effet utile sur la cognition

ou sur le contrôle, mais qui, pour des raisons darwiniennes usuelles, ne se trouvent pas éliminées et qui peuvent même dériver vers une fixation (culturelle *ou* génétique) dans certaines sous-populations. Par exemple : se peindre en bleu, se fouetter avec des rameaux, se scarifier, jeûner, se répéter sans cesse une formule « magique », et contempler son nombril. Si ces pratiques sont des habitudes qui valent la peine qu'on se les inculque, leurs vertus en tant que susceptibles de développer la *fitness* ne sont, c'est le moins qu'on puisse dire, pas assez « évidentes » pour qu'elles se soient propulsées sous la forme de prédispositions génétiques. Mais il se peut que ce soient des inventions trop récentes.

Les diverses formes d'autostimulation qui permettent le développement de l'organisation cognitive sont aujourd'hui probablement en partie innées et en partie apprises et idiosyncrasiques. Tout comme on peut remarquer que le fait de se frapper de certaines manières peut produire certains effets secondaires désirables qui ne sont contrôlables que partiellement et indirectement – on peut ensuite consacrer un peu de temps et d'énergie pour développer et explorer les techniques propres à produire ces effets secondaires – en sorte que l'on peut, semi-consciemment, explorer des techniques d'autostimulation cognitive, en développant un style personnel pourvu de certaines forces et de certaines faiblesses. Certaines personnes sont meilleures que d'autres pour ce genre de tâches, et certaines n'apprendront jamais les techniques nécessaires, mais on peut partager et enseigner beaucoup de choses. La *transmission* culturelle, en laissant presque tout le monde accéder au Bon Truc, peut aplatir le dessus de la colline de la *fitness* (voir figure 6.2, p. 187), en créant une butte ou un plateau qui diminue la pression sélective de manière à mouvoir le Truc dans le génome. Si presque tout le monde devient assez bon pour s'insérer dans le monde civilisé, la pression sélective qui envoie les Bons Trucs dans le génome s'éteint ou s'en trouve du moins diminuée.

Le troisième processus évolutif : les mêmes et l'évolution culturelle ¹⁴

Nous apprenons à traire les vaches, puis à les domestiquer pour notre bénéfice ; de même nous apprenons à traire les autres

et notre propre esprit de certaines façons. À présent, les techniques d'autostimulation mutuelle sont profondément ancrées dans notre culture et dans notre apprentissage. Pour comprendre la manière dont s'est construite la conscience humaine, qui est un véhicule de l'évolution, il est important de comprendre comment la culture est devenue le dépositaire et le véhicule de la transmission des innovations (et pas seulement de celles de la conscience).

L'une des étapes majeures franchies par le cerveau humain dans le processus massif de l'auto-organisation postnatale est l'ajustement aux conditions locales les plus importantes : il se transforme rapidement (en deux ou trois ans) en un cerveau swahili, japonais ou anglais ! Quelle étape – franchie presque comme d'un coup de lance-pierre !

Pour ce qui nous occupe ici, appeler ce processus « apprentissage » ou « développement différentiel » importe peu : il se produit si rapidement et sans effort qu'il est presque certain que le génotype humain inclut de nombreuses adaptations qui sont spécifiquement en place pour développer l'acquisition du langage. Tout cela s'est produit très rapidement, en termes d'évolution. On pouvait s'y attendre, étant donné l'effet Baldwin. Être capable de parler est un tellement Bon Truc que quiconque aurait mis du temps à l'acquérir aurait eu un désavantage énorme. Les premiers de nos ancêtres qui ont parlé ont conquis cette faculté certainement de manière bien plus laborieuse, mais nous sommes les descendants des virtuoses du groupe ¹⁵.

À partir du moment où notre cerveau a bâti les voies d'entrées et de sortie des véhicules du langage, il devient rapidement *parasité* (au sens littéral de ce terme, comme nous allons le voir) par des entités qui ont évolué pour lutter précisément dans ce genre de niches : les *mêmes*. Les grandes lignes de la théorie de l'évolution par sélection naturelle sont claires : l'évolution survient chaque fois que les conditions suivantes existent :

1. variation : un afflux continu de différents éléments ;
2. hérédité ou réplique : les éléments ont la capacité de créer des copies ou des répliques d'eux-mêmes ;
3. *fitness* différentielle : le nombre des copies d'un élément qui sont créées à un moment donné, selon les interactions entre les traits de cet élément (quel que soit ce qui le rend différent des autres éléments) et les traits de l'environnement dans lesquels il persiste.

Notez que cette définition, bien qu'elle soit tirée de la biologie, ne dit rien de spécifique sur les molécules organiques, la nutrition ou même la vie. C'est une caractérisation plus générale et abstraite de l'évolution par sélection naturelle. Comme l'a remarqué le zoologue Richard Dawkins, le principe fondamental est

« que toute vie évolue par survie différentielle d'entités qui se répliquent...

Le gène, la molécule d'ADN, se trouve être l'entité de réplication qui domine sur notre planète. Il peut y en avoir d'autres. S'il y en a, et si certaines autres conditions sont présentes, ils tendront presque inévitablement à devenir la base du processus d'évolution.

Mais devons-nous aller vers des mondes distants pour découvrir d'autres sortes de réplication et par conséquent d'autres types d'évolution ? Je pense qu'un nouveau type de réplicateur vient récemment d'émerger sur cette planète. Il nous regarde en face. Il est toujours dans son enfance, dérivant maladroitement dans sa soupe primitive, mais il est déjà en train de produire un changement évolutif à une vitesse telle qu'il laisse l'ancien gène à bonne distance » (1976, p. 206).

Ces nouveaux réplicateurs sont, en gros, les idées. Non pas les « idées simples » de Locke et de Hume (l'idée du rouge, ou l'idée de rondeur ou celle de chaud ou de froid), mais les idées complexes qui se concentrent en unités mémorables. Ce sont par exemple les idées suivantes :

roue
porter des habits
vendetta
triangle rectangle
alphabet
calendrier
Odyssée
calcul différentiel
échecs
dessein de la perspective
évolution par sélection naturelle
impressionnisme
« Marseillaise »
le déconstructionnisme

De prime abord, ce sont des unités culturelles plus ou moins identifiables, mais nous pouvons être plus précis quant à la manière dont nous traçons les frontières. Pourquoi, par exemple

ré fa-dièse-la n'est-il pas une unité et pourquoi le thème du mouvement lent de la *Septième Symphonie* de Beethoven en est-il une ? Les unités sont les plus petits éléments susceptibles de se répliquer de façon fiable et féconde. Dawkins invente un terme pour ce genre d'unités : les *mêmes*. Un même est

une unité de transmission culturelle, ou une unité d'*imitation* « Mimème » vient d'une racine grecque appropriée, mais je veux un monosyllabe qui sonne un peu comme « gène »... et qu'on pourrait aussi penser comme relié à « mémoire » ou au mot français *même*...

Des exemples de mêmes sont : des airs musicaux, des idées, des expressions toutes faites, des modes vestimentaires, des façons de faire des poteries ou de construire des voûtes. Tout comme les gènes se propagent dans le pool génétique en sautant d'un corps à un autre par l'intermédiaire du sperme et des œufs, les mêmes se propagent dans le pool mémétique en sautant de cerveau en cerveau à travers un processus qui, au sens large, peut être appelé imitation. Si un savant entend, ou lit quelque part, une bonne idée, il la passe à ses collègues et à ses étudiants. Il la mentionne dans ses articles et dans ses conférences. Si l'idée retient l'attention, on peut dire qu'elle se propage, se répandant de cerveau en cerveau (1976, p. 206).

Dans *Le Gène égoïste*, Dawkins nous demande de prendre l'idée d'évolution des mêmes au sens littéral. L'évolution des mêmes n'est pas seulement analogue à l'évolution biologique des gènes ; ce n'est pas seulement un processus que l'on peut décrire métaphoriquement dans ces idiomes évolutionnistes, mais un phénomène qui obéit exactement aux lois de la sélection naturelle. La théorie de l'évolution par sélection naturelle est neutre par rapport aux différences entre les mêmes et les gènes ; ce sont seulement deux types différents de réplicateurs qui évoluent dans des milieux différents à des vitesses différentes. Tout comme les gènes pour les animaux ne pouvaient pas apparaître sur cette planète tant que l'évolution des plantes n'avait pas ouvert la voie (en créant l'atmosphère riche en oxygène et un stock disponible de nourritures convertibles), l'évolution des mêmes n'a pu démarquer tant que l'évolution des animaux n'a pas ouvert la voie en créant une espèce – l'*Homo sapiens* – dotée d'un cerveau capable de fournir des abris et des habitudes de communication qui puissent produire des véhicules de transmission pour les mêmes.

C'est une nouvelle façon de concevoir les idées. C'est aussi, j'espère le montrer, une bonne façon de les concevoir. Mais de

prime abord la perspective qu'elle nous fournit est dérangeante, voire même choquante. Nous pouvons la résumer par un slogan :

un chercheur n'est autre que la façon dont une bibliothèque crée une autre bibliothèque.

Je ne sais pas ce qu'il en est pour vous, mais je n'éprouve pas d'attirance immédiate pour l'idée que mon cerveau serait une sorte de tas de fumier dans lequel les larves des idées d'autres gens se renouvellent, avant d'envoyer leurs copies dans une diaspora informationnelle. Cela semble ôter à mon esprit son importance comme auteur et comme critique. Qui commande, selon cette façon de voir les choses, nous ou nos mêmes ?

Bien entendu, il n'est pas facile de répondre. C'est ce qui explique les confusions qui entourent l'idée de *soi*. La conscience humaine est dans une large mesure le produit non seulement de la sélection naturelle, mais aussi de l'évolution culturelle. Pour évaluer la contribution des mêmes à la création de nos esprits, le mieux est de suivre les étapes ordinaires de la pensée évolutionniste de plus près.

La première règle des mêmes, comme c'est le cas pour les gènes, est que la réplication n'est pas nécessaire pour qu'un dispositif soit bon à quelque chose ; certains répliqueurs fleurissent du fait même qu'ils sont bons... à se répliquer, et ce quelle que soit la raison pour laquelle ils le font. Comme l'a dit Dawkins :

« Un même qui a fait sauter ses corps du haut des falaises aurait un destin semblable à celui d'un gène qui ferait sauter les corps du haut des falaises. Il tendrait à être éliminé du pool mémétique... Mais cela ne veut pas dire que le critère du succès de la sélection des mêmes ne soit pas en dernière instance la survie des gènes... Il est évident qu'un même qui conduit les individus qui le portent à se tuer a un grave désavantage, mais celui-ci n'est pas nécessairement fatal... Un même suicidaire peut se répandre, comme quand un martyr dramatique et bien publicisé inspire aux autres le désir de mourir pour une cause profondément chérie, ce qui en retour inspire les autres à mourir, et ainsi de suite » (1982, p. 110-111).

Il est important de noter qu'il n'existe pas de connexion *nécessaire* entre le pouvoir réplcatif d'un même, sa *fitness* de son point de vue, et sa contribution à *notre fitness* (quels que soient nos critères pour en juger). La situation n'est pas totalement désespérée. Alors que certains mêmes, de toute évidence, nous

manipulent pour que nous collaborions à leur répllication *en dépit* du fait que nous les jugions inutiles, horribles ou même dangereux pour notre santé et pour notre bien-être, beaucoup de mêmes – la plupart d’entre eux, si nous avons de la chance – qui se répliquent ne le font pas seulement *avec* notre bénédiction, mais *à cause de* l’estime que nous leur portons. Je pense que l’on ne contestera pas le fait que certains mêmes sont, tout bien considéré, bons *de notre point de vue*, et non pas de leur propre point de vue, en tant que répliqueurs égoïstes d’eux-mêmes : des mêmes généraux tels que la coopération, la musique, l’écriture, l’éducation, la conscience de l’environnement, la réduction des armements, et des mêmes particuliers tels que *Le mariage de Figaro*, *Moby Dick*, consigner les bouteilles, les accords SALT. D’autres mêmes sont plus contestables ; nous pouvons voir pourquoi ils se répandent, et pourquoi, tout bien considéré, nous devrions les tolérer, en dépit des problèmes qu’ils causent pour nous : les centres commerciaux, la restauration rapide, la publicité télévisée. D’autres encore sont sans conteste pernicioeux, mais très difficiles à éradiquer : l’antisémitisme, les détournements d’avions, les virus informatiques, les tags et les graffitis.

Les gènes sont invisibles ; ils sont transportés par des véhicules à gènes (les organismes) dans lesquels ils tendent à produire des effets (des effets « phénotypiques ») qui déterminent, à long terme, leur destin. Les mêmes sont également invisibles, ils sont transportés par des véhicules à mêmes – des images, des livres, des déclarations (dans des langages particuliers, oraux ou écrits, sur papier ou encodés magnétiquement, etc.). Les outils, les buildings et autres inventions servent aussi à véhiculer des mêmes. Un wagon à roues ne transporte pas seulement du grain ou du fret d’un lieu à un autre : il transporte l’idée brillante d’un wagon à roues d’un esprit à l’autre. L’existence d’un même dépend de son incarnation physique dans un véhicule quelconque : si toutes ces incarnations physiques sont détruites, ce même s’éteint. Il peut, bien entendu, faire ensuite indépendamment sa réapparition – tout comme les gènes de dinosaures pourraient, en principe, se réassembler dans un lointain futur – mais les dinosaures qu’ils créeraient et qu’ils habiteraient ne seraient pas les descendants des dinosaures initiaux – ou tout au moins pas plus directement que nous ne sommes nous-mêmes leurs descendants. Le destin des mêmes – le fait que leurs copies et les copies de leurs copies se multiplient – dépend des forces sélectives qui

agissent directement sur les véhicules physiques qui les incarnent.

Les véhicules des mêmes habitent notre monde en même temps que toute la faune et la flore, grande ou petite. Dans l'ensemble, ils ne sont cependant « visibles » qu'aux yeux de l'espèce humaine. Considérons l'environnement du pigeon parisien moyen, dont les yeux et les oreilles subissent quotidiennement l'assaut d'autant de mots, d'images et d'autres signes et symboles que chaque Parisien humain. Ces véhicules physiques de mêmes peuvent peser de façon importante sur le bien-être du pigeon, mais pas en vertu de mêmes qu'ils transportent – pour le pigeon qui picore, peu importe qu'il trouve une miette sur *France-Soir* ou sur *Le Monde*. Pour les êtres humains, en revanche, chaque véhicule de mêmes est un ami ou un ennemi potentiel, porteur d'un cadeau qui développera nos pouvoirs ou d'un cheval qui nous distraira, alourdira notre mémoire, dérangera notre jugement. Nous pouvons comparer ces envahisseurs de nos yeux et de nos oreilles suspendus dans les airs aux parasites qui entrent dans notre corps par d'autres voies : certains sont bénéfiques, comme les bactéries de notre système digestif sans lesquelles nous ne digérerions pas notre nourriture ; d'autres sont tolérables et ne valent pas la peine que nous les éliminions (tous les habitants de notre peau et de notre cuir chevelu, par exemple) ; et les envahisseurs pernicioeux qui sont difficiles à éradiquer (le virus du SIDA, par exemple).

Jusqu'ici, adopter le point de vue du même semblera simplement être une manière commode d'organiser des observations familières quant à la manière dont des entités culturelles nous affectent et s'affectent les unes les autres. Mais Dawkins suggère que dans nos explications nous tendons à négliger le fait fondamental qu'« un trait culturel peut avoir évolué de la manière dont il a évolué simplement parce qu'il est *avantageux pour lui-même* » (1976, p. 214). C'est ce qui permet de répondre à la question de savoir si oui ou non le même fait partie de ceux que nous devrions exploiter et répliquer. Selon la façon de voir normale, les énoncés suivants sont pratiquement tautologiques :

Les gens ont cru à l'idée X parce que X était considérée comme vraie. Les gens ont approuvé X parce qu'ils ont trouvé X superbe.

Ce qui requiert une explication spécifique, ce sont les cas dans lesquels, en dépit de la vérité et de la beauté d'une idée, celle-ci n'est *pas* acceptée, ou ceux dans lesquels, en dépit de son

caractère repoussant ou de sa fausseté, elle *est* acceptée. Le point de vue du même vise à être une perspective possible à partir de laquelle on peut expliquer ces déviations. Ce qui est tautologique *de son point de vue*, c'est :

Que le même X se soit répandu parmi la population parce que X était un bon réplicateur.

Or il existe une corrélation non aléatoire entre les deux : ce n'est pas un accident. Nous ne pourrions pas survivre si nous avions seulement pris l'habitude de choisir au hasard les mêmes qui nous sont utiles. Nos systèmes immunologiques-à-mêmes ne sont pas totalement à toute épreuve, mais ils ne sont pas inutiles pour autant. Nous pouvons compter, en général et en gros, sur le fait que les deux perspectives coïncident : dans l'ensemble, les bons mêmes sont ceux qui sont aussi les bons réplicateurs.

La théorie ne devient intéressante que quand nous envisageons les exceptions, les circonstances dans lesquelles les deux perspectives divergent ; ce n'est que si la théorie des mêmes nous permet de comprendre mieux les déviations à partir du schème normal qu'elle sera susceptible d'être acceptée. (Notez que dans ses propres termes, le fait que le même des mêmes se réplique avec succès est strictement indépendant de sa vertu épistémologique : il pourrait se répandre en dépit de son caractère pernicieux, ou s'éteindre en dépit de sa vertu.)

Les mêmes se répandent aujourd'hui à travers le monde à la vitesse de la lumière ; ils se répliquent à des vitesses qui en comparaison font paraître glaciales les mouches à fruit et les cellules de levure. Ils sautent, dans une complète promiscuité, d'un véhicule à l'autre, d'un milieu à l'autre, et se révèlent virtuellement impossibles à mettre en quarantaine. Les mêmes, comme les gènes, sont *potentiellement* immortels, mais, comme les gènes ils dépendent de l'existence d'une chaîne continue de véhicules physiques qui résiste à la seconde loi de la thermodynamique. Les livres sont relativement permanents, et les inscriptions sur les monuments le sont encore plus. Mais à moins qu'ils ne soient sous la protection de conservateurs humains, ils tendent à se dissoudre au cours du temps. Comme dans le cas des gènes, l'immortalité est plus affaire de réplication que de longévité des véhicules individuels. La préservation des mêmes platoniciens, à travers une série de copies de copies, en est un

exemple particulièrement frappant. Bien qu'on ait découvert récemment certains fragments de papyrus des textes de Platon qui sont en gros contemporains de Platon lui-même, la survie des mêmes ne doit presque rien à une persistance à long terme de ce genre. Les bibliothèques actuelles contiennent des milliers, voire des millions d'exemplaires physiques (et de traductions) du *Ménon* de Platon, et les ancêtres clefs dans la transmission de ce texte sont tombés en poussière il y a des siècles.

La réplique physique brute des véhicules ne suffit pas à assurer la longévité des mêmes. Quelques milliers d'exemplaires reliés d'un nouveau livre peuvent disparaître en quelques années sans quasiment laisser de traces ; qui sait combien de brillantes lettres adressées aux journaux, reproduites à des centaines de milliers d'exemplaires, disparaissent dans des décharges et des incinérateurs chaque jour ? Un jour viendra peut-être où des évaluateurs non humains de mêmes suffiront pour sélectionner et assurer la préservation de mêmes particuliers. Mais pour l'instant les mêmes dépendent au moins de façon indirecte du fait qu'un ou plusieurs de leurs véhicules, telles des chrysalides, passent au moins un bref séjour dans un nid à mêmes de type remarquable : un esprit humain.

Le stock d'esprits disponibles est fini, et chaque esprit a une capacité limitée à acquérir des mêmes. Par conséquent, la compétition est considérable entre les mêmes pour se reproduire sous autant de types que possible. Cette compétition est la force sélective majeure dans la mémosphère, et, tout comme dans la biosphère, le défi a été relevé avec une grande ingéniosité. Par exemple, quelles que soient les vertus caractérisant les mêmes suivants, ils ont en commun la propriété d'avoir des expressions phénotypiques qui tendent à rendre leur propre réplique plus probable en affaiblissant ou en contrecarrant les forces environnementales qui tendent à les faire s'éteindre : le même de la *foi*, qui décourage l'exercice du jugement critique qui pourrait décider que l'idée de foi est tout bien considéré une idée dangereuse (Dawkins, 1976, p. 212) ; le même de la *tolérance* ou de la *liberté d'expression* ; le même d'inclusion dans une chaîne de lettres d'un avertissement au sujet de la destinée terrible de ceux qui ont brisé la chaîne dans le passé ; le même de la *théorie de la conspiration*, qui a une réponse toute prête à l'objection selon laquelle il n'y a pas de preuve que la conspiration existe : « Bien sûr qu'il n'y en a pas — cela prouve à quel point la conspiration est puissante ! » Certains

de ces mêmes sont « bons » et d'autres peuvent être « mauvais » ; ce qu'ils ont en commun, c'est un effet phénotypique qui tend systématiquement à désarmer les forces sélectives dirigées contre eux. Toutes choses égales par ailleurs, la mémétique des populations prédit que les mêmes de la théorie de la conspiration persisteront indépendamment de leur vérité ou de leur fausseté, et le même de la foi est propre à assurer sa propre survie, ainsi que celui des mêmes religieux qui le chevauchent, même dans les environnements les plus rationalistes. En fait, le même de la foi manifeste une *fitness-relative-à-la-fréquence* : il fleurit d'autant mieux qu'il est surpassé en nombre par des mêmes rationalistes ; dans un environnement où on trouve peu de sceptiques, le même de la foi tend à s'estomper et à ne plus être utilisé.

D'autres concepts appartenant à la génétique des populations se laissent aussi transférer aisément. Voici un cas de ce qu'un généticien appellerait des *lieux liés* : deux mêmes qui se trouvent physiquement liés l'un à l'autre en sorte qu'ils tendent toujours à se répliquer ensemble, ce qui affecte leurs chances. Il existe une magnifique marche cérémoniale, familière à beaucoup d'entre nous, qui serait particulièrement très appropriée pour les inaugurations, les mariages et d'autres fêtes ; elle pourrait faire disparaître *Pompe et circonstance* et la Marche nuptiale de *Lohen-grin*. Seulement son même musical est trop étroitement lié au même de son titre, de sorte que beaucoup d'entre nous tendent à penser à lui chaque fois que nous entendons l'Exécuteur des Hautes Œuvres : l'inusable chef-d'œuvre d'Arthur Sullivan « Gloire à Son Excellence » (*Behold the Lord High Executioner*).

Pour tous les mêmes, le port qu'il faut atteindre est l'esprit humain, mais un esprit humain est lui-même un artefact créé quand les mêmes restructurent un cerveau humain afin d'en faire un meilleur habitat pour les mêmes. Les avenues d'entrée et de départ sont modifiées pour répondre aux conditions locales ; elles sont renforcées par divers dispositifs artificiels qui développent la fidélité et la prolixité de la répllication : les esprits qui sont nés chinois diffèrent de façon radicale des esprits français, et les esprits cultivés diffèrent des esprits incultes. Ce que les mêmes produisent en retour pour les organismes dans lesquels ils résident, c'est un magasin incalculable d'avantages parmi lesquels on trouve sans doute quelques chevaux de Troie, pour faire bonne mesure. Les cerveaux humains normaux ne sont pas tous identiques : ils varient considérablement par leur taille, par

leur forme et par les myriades de détails que comportent les connexions sur lesquelles reposent leurs prouesses. Mais les différences les plus frappantes entre les prouesses humaines dépendent de différences microstructurales induites par les divers mêmes qui y sont entrés et qui y ont pris résidence. Les mêmes améliorent leurs possibilités respectives : le même de l'éducation, par exemple, est un même qui renforce le processus d'implantation de mêmes.

Mais s'il est vrai que les esprits humains sont eux-mêmes dans une très large mesure les créations des mêmes, nous ne pouvons soutenir la polarité de vision dont nous sommes partis ; cela ne peut pas être une affaire de « mêmes contre nous », parce que les invasions antérieures de mêmes ont déjà joué un rôle majeur dans la détermination de *qui nous sommes ou de ce que nous sommes*. L'esprit « indépendant » qui combat pour se protéger des mêmes étrangers et dangereux est un mythe ; en sous-sol, la tension persiste entre l'impératif biologique des gènes et celui des mêmes, mais nous serions stupides de « prendre le parti » de nos gènes : ce serait commettre l'erreur la plus fameuse de la pop-sociobiologie. Sur quel fondement, par conséquent, pouvons-nous nous reposer quand nous nous battons pour garder pied dans le mèmestrom dans lequel nous sommes engouffrés ? Si ce qui est possible répliquativement (*might*) ne signifie pas « correctement » (*right*), en quoi consistera l'idéal relativement auquel « nous » jugerons la valeur des mêmes ? Nous devrions noter que les mêmes responsables des concepts normatifs – pour *devrait*, *bon* et pour *vérité* et *beauté* – sont au nombre des habitants les plus enracinés de nos esprits, et que parmi les mêmes qui nous constituent, ils jouent un rôle central. Notre existence en tant que nous-mêmes, en tant que ce que nous sommes comme êtres pensants – et non pas en tant que ce que nous sommes comme organismes – n'est pas indépendante de ces mêmes.

Pour résumer : l'évolution des mêmes a le potentiel nécessaire pour contribuer au développement remarquable de l'organisation de la machinerie sous-jacente du cerveau – à une grande vitesse, si on la compare au rythme lent de la recherche et de la découverte. L'idée lamarckienne, discréditée, selon laquelle les caractères acquis se transmettaient de façon génétique attirait les biologistes au départ parce qu'elle avait la capacité présumée d'accélérer de nouvelles inventions dans le génome. (Pour une belle réfutation du lamarckisme, voir la discussion détaillée de Dawkins dans *The*

Extended Phenotype, 1982.) Cela n'arrive pas et ne peut arriver. L'effet Baldwin n'accélère pas l'évolution en favorisant le mouvement de Bons Trucs découverts par des individus en direction du génome, à travers un chemin indirect, en créant de nouvelles pressions sélectives résultant de l'adoption par une grande quantité d'individus des Bons Trucs. Mais l'évolution culturelle, qui a un rythme beaucoup plus rapide, permet aux individus d'acquérir, par l'intermédiaire de la transmission culturelle, les Bons Trucs qu'ont abrités des prédécesseurs qui ne sont même pas leurs ancêtres génétiques. Les effets de ce partage des bonnes organisations sont si puissants que l'évolution culturelle a probablement oblitéré toutes les pressions douces de l'effet Baldwin, à quelques exceptions près. Les améliorations d'organisation que l'on reçoit de sa culture – on a rarement à « réinventer la roue » – submergent sans doute la plupart des différences génétiques individuelles quant à l'architecture du cerveau, ôtant les avantages que peuvent acquérir ceux qui sont un petit peu mieux lotis à la naissance.

Ces trois véhicules – l'évolution génétique, la plasticité phénotypique, et l'évolution mémétique – ont contribué à l'organisation de la conscience humaine, chacun à son tour, et à des vitesses de développement sans cesse croissantes. Comparée à la plasticité phénotypique, qui opère sans doute depuis des millions d'années, l'évolution mémétique significative est un phénomène extrêmement récent, qui n'est devenu une force puissante que durant les cent mille dernières années et qui a explosé avec le développement de la civilisation il y a moins de dix mille ans. Elle se limite à une seule espèce, l'*Homo sapiens*, et nous pourrions noter qu'elle nous a désormais conduits à l'aube d'un quatrième véhicule de potentiel R et D, grâce aux mêmes de la science : la révision directe des systèmes nerveux individuels par ingénierie neuroscientifique et la révision du génome par ingénierie génétique.

Les mêmes de la conscience : la machine virtuelle à installer

« Bien qu'un organe ait pu ne pas avoir été originellement formé pour un but spécifique quelconque, s'il sert à présent cet objectif nous avons le droit de dire qu'il est spécifiquement construit pour lui. Selon le même principe, si un homme devait faire une machine pour un but spécifique, mais devait

utiliser des vieilles roues, ressorts et poulies, seulement légèrement modifiés, la machine tout entière, avec ses parties, pourrait être dite avoir été spécifiquement construite pour cet objectif. Ainsi à travers toute la nature presque toute partie d'être vivant a probablement servi, sous une forme légèrement modifiée, à divers objectifs, et a agi dans la machinerie vivante de nombreuses formes anciennes et distinctes. »

Charles Darwin, 1874.

« Le grand cerveau, comme un grand gouvernement, peut ne pas être capable de faire des choses simples de façon simple. »

Donald Hebb, 1958.

« L'instinct le plus puissant dans le développement humain est son plaisir à prouver sa propre habileté. Il aime faire ce qu'il fait bien, et une fois qu'il l'a fait bien il aime à la faire mieux. »

Jacob Bronowski, 1973.

Mon récit spéculatif met en évidence le fait que nos ancêtres, comme nous, ont pris plaisir à divers modes d'exploration de soi relativement indirects – se stimuler soi-même sans cesse et voir ce qui se passe. En raison de la plasticité du cerveau, couplée à l'inquiétude constante et à la curiosité qui nous conduit à explorer tout coin et recoin de notre environnement (dont nos corps sont un élément si important et si permanent), il n'est pas surprenant, à la réflexion, que nous tombions sur des stratégies d'autostimulation ou d'automanipulation ; celles-ci ont conduit à l'inculcation d'habitudes et de dispositions qui ont altéré radicalement la structure communicative interne de nos cerveaux. On ne doit pas non plus s'étonner que ces découvertes soient devenues une partie de notre culture – les mêmes – qui deviennent ainsi disponibles pour tous.

La transformation d'un cerveau humain par des invasions de mêmes est une altération majeure de la compétence de cet organe. Comme nous l'avons noté, les différences dans un cerveau dont le langage maternel est le chinois plutôt que l'anglais rendraient compte d'énormes différences dans la compétence de ce cerveau, qui seraient reconnaissables instantanément dans le comporte-

ment et qui se révéleraient dans de nombreux contextes. Rappelez-vous, par exemple, combien il est important, pour l'expérimentateur (l'hétérophénoménologue), dans les expériences avec les sujets humains, de savoir si les sujets ont compris les instructions. Ces différences fonctionnelles, bien qu'elles soient sans doute toutes incarnées dans des structures de changements microscopiques dans le cerveau, sont aussi bénéfiques qu'invisibles aux yeux des neurobiologistes, aujourd'hui et probablement à jamais. Par conséquent, si nous voulons accéder à l'architecture fonctionnelle *créée* par ces manifestations de mêmes, il nous faudra trouver un niveau supérieur pour le décrire. Par chance, on en a un, que nous tirons de l'informatique. Le niveau de description et d'explication dont nous avons besoin est *analogue* à (mais pas identique à) la description des « niveaux de logiciel » pour les ordinateurs : nous devons comprendre comment la conscience humaine peut être réalisée dans l'opération d'une *machine virtuelle* créée par des mêmes dans le cerveau.

Je veux défendre l'hypothèse suivante :

La conscience humaine est *elle-même* un énorme complexe de mêmes (ou plus exactement d'effets de mêmes dans le cerveau) ; elle fonctionne en quelque sorte comme une machine virtuelle à la von Neumann *implémentée* dans l'*architecture parallèle* d'un cerveau, lequel n'était pas conçu pour de telles activités. Les pouvoirs de cette *machine virtuelle* augmentent énormément les pouvoirs sous-jacents du *matériel* organique sur lequel elle tourne ; en même temps, beaucoup de ses propriétés les plus curieuses, et en particulier ses limitations, peuvent s'expliquer comme des sous-produits des *kludges* rendant possible cette réutilisation curieuse mais effective d'un organe déjà existant en vue d'objectifs nouveaux.

Le sens de cette hypothèse se dégagera bientôt de la gangue de jargon dont elle est entourée dans la formulation qui vient d'être donnée. Pourquoi ai-je utilisé ce jargon ? Parce que ce sont des termes désignant des concepts importants dont les théoriciens de l'esprit ne disposent que depuis peu de temps. Il n'existe pas d'autres mots qui puissent exprimer ces concepts de façon nette, et ils méritent qu'on les présente. C'est précisément ce que je vais faire, en m'aidant d'une brève digression historique et en plaçant ces concepts dans le contexte dans lequel nous allons les utiliser ensuite.

Deux des inventeurs les plus importants de l'ordinateur furent le mathématicien britannique Alan Turing et le mathématicien

et physicien hongrois John von Neumann. Bien que Turing se soit beaucoup mêlé d'expériences pratiques qui portaient sur la conception et la construction de machines électroniques servant à déchiffrer les codes et qui ont contribué à la victoire des Alliés pendant la Seconde Guerre mondiale, c'est son travail purement abstrait et théorique, en particulier sa conception de la *Machine de Turing Universelle*, qui a ouvert l'âge des ordinateurs. Von Neumann, quant à lui, a compris comment utiliser le schème abstrait de Turing (il était en fait « philosophique » : c'était une expérience de pensée et non une conception d'ingénierie) et comment le rendre assez concret pour en faire le modèle (toujours abstrait) d'un ordinateur électronique bien réel et utilisable en pratique. Ce modèle abstrait, connu sous le nom d'*architecture de von Neumann*, est celui que l'on peut trouver aujourd'hui dans presque tout ordinateur au monde, des « ordinateurs centraux » aux puces placées à l'intérieur du plus modeste des ordinateurs personnels.

Un ordinateur possède une architecture de base *fixe* ou câblée au départ, mais il a aussi une grande plasticité en raison de sa *mémoire* ; celle-ci peut stocker à la fois des programmes (connus sous le nom de *logiciels*) et des *données*, qui sont les trames seulement transitoires suivant à la trace tout ce qui doit être représenté. À leur naissance, les ordinateurs, comme les cerveaux, sont incomplets ; mais ils sont dotés d'une flexibilité qui peut servir de véhicule permettant de créer un plus grand nombre d'architectures plus disciplinées, c'est-à-dire de machines à tâches spécifiques ; chacune possède alors sa manière propre de saisir les stimuli de l'environnement (par l'intermédiaire du clavier ou d'autres dispositifs d'entrée d'information), pour produire des réponses (par un écran cathodique ou par d'autres dispositifs de sortie d'information).

Ces structures temporaires sont « faites de règles plutôt que de câbles » : les informaticiens les appellent des *machines virtuelles* ¹⁶. Vous obtenez une machine quand vous imposez une structure particulière de règles (plus littéralement : des dispositions ou des régularités de transition) sur toute cette plasticité. Considérons le cas de quelqu'un qui s'est cassé le bras et à qui on a mis un plâtre. Le plâtre restreint sérieusement le mouvement de son bras ; son poids et sa forme induisent également la personne à ajuster le reste de ses mouvements corporels. Considérons à présent un mime (par exemple, Marcel Marceau) imitant

quelqu'un qui a un plâtre au bras : si le mime se débrouille bien, ses mouvements corporels sont restreints à peu près exactement de la même façon. Il a alors un *plâtre virtuel* sur son bras – et celui-ci est « presque visible ». Quiconque a l'habitude d'utiliser un traitement de texte connaît au moins une machine virtuelle, et si vous avez utilisé différents traitements de texte, utilisé un tableur ou joué à un jeu sur le même ordinateur que celui que vous utilisez pour du traitement de texte, vous connaissez plusieurs machines virtuelles, qui reposent sur les actions d'une machine réelle particulière. Les différences deviennent tout à fait visibles, en sorte que l'utilisateur sait à tout moment avec quelle machine virtuelle il entre en interaction.

Tout le monde sait que différents programmes donnent aux ordinateurs différents pouvoirs, mais tout le monde ne connaît pas les détails. Certains d'entre eux sont importants pour notre récit. Je vous demande donc un peu d'indulgence. Un bref résumé du processus inventé par Alan Turing est indispensable.

Turing n'essayait pas d'inventer le traitement de texte ou le jeu vidéo quand il a fait ses superbes découvertes. Il réfléchissait, consciemment et introspectivement, à la manière dont lui-même, mathématicien, résolvait des problèmes mathématiques ou accomplissait des *calculs* ; il a fait un pas décisif en essayant de décomposer la *suite* de ses *actes mentaux* en éléments de base. « Qu'est-ce que je fais, s'est-il sans doute demandé, quand je fais un calcul ? Eh bien, je me demande tout d'abord quelle est la règle à appliquer, et quand j'applique la règle, puis j'écris le résultat, puis je regarde le résultat, puis je me demande ce que je dois faire ensuite, etc. » Turing était un penseur extraordinairement bien organisé, mais son flux de conscience, comme le vôtre, le mien ou celui de James Joyce, était sans aucun doute un chaos d'images variées, de décisions, d'intuitions, de souvenirs, etc., à partir desquelles il réussissait à distiller la quintessence mathématique : le squelette, la séquence minimale d'opérations permettant de parvenir à l'objectif qu'il se fixait au milieu des activités pleines de méandres dans lesquelles s'absorbait son esprit conscient. Le résultat fut la spécification de ce que nous appelons aujourd'hui une machine de Turing, qui est une idéalisation brillante et une simplification d'un phénomène hyperrationnel et hyperintellectuel : celui du mathématicien effectuant un calcul rigoureux. L'idée de base contenait cinq éléments :

1. un processus *sériel* (des événements se produisant un par un), dans
2. un *espace de travail* sérieusement réduit, auquel
3. on apportait à la fois des *données* et des *instructions*,
4. provenant d'une *mémoire* inerte mais extrêmement fiable,
5. qui devaient être effectuées à partir d'un ensemble fini d'*opérations primitives*.

Dans la formulation initiale de Turing, l'espace de travail était conçu comme un scanner examinant une seule case à la fois sur la bande de papier, pour voir si un zéro ou un un était écrit dessus. Selon ce qu'elle « voyait », ou bien elle effaçait le zéro ou le un et imprimait l'autre symbole, ou bien elle laissait la case inchangée. Elle déplaçait alors la bande vers la gauche ou vers la droite d'une case et regardait ensuite ; dans chaque cas, elle était gouvernée par un ensemble fini d'instructions programmées au départ et formant sa *table de machine*. La bande constituait la mémoire.

L'ensemble d'opérations primitives de Turing (les actes « considérés comme atomiques par l'introspection ») étaient délibérément appauvries, en sorte que la possibilité de les réaliser mécaniquement ne posait pas de problème. Cela veut dire qu'il importait beaucoup à Turing, mathématiquement parlant, qu'il n'y ait aucun doute sur le fait que chacune des étapes dans le processus qu'il étudiait soit une étape si simple et si stupide qu'il puisse être accompli par un simpleton, par un individu qu'une machine pourrait remplacer : REGARDER, EFFACER, IMPRIMER, AVANCER DE GAUCHE À DROITE, et ainsi de suite.

Turing savait que cette spécification idéale pourrait servir, indirectement, de moule permettant de construire une machine à calculer effective. D'autres l'ont compris aussi, en particulier John von Neumann, qui a modifié les idées de base de Turing et a ainsi créé l'architecture abstraite du premier ordinateur digital réalisable en pratique. Nous pouvons appeler cette architecture la *machine de von Neumann*.

Sur la gauche, on trouve la mémoire ou *random access memory* (RAM), où l'on conserve à la fois les données et les instructions, codées sous forme de suites de chiffres binaires ou « bits », tels que 00011011 ou 01001110. Le processus sériel de Turing a lieu dans l'espace de travail composé de deux « registres » appelés *accumulateur* et *registre à instructions*. Une instruction est copiée

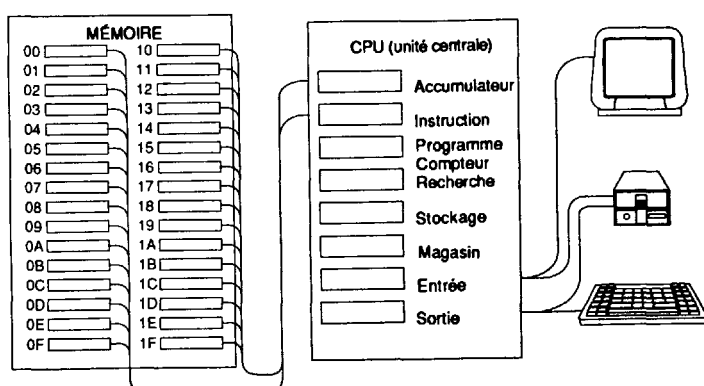


Figure 6.4

De Macrocomputer, par Steven Barney, © Tufts Curricular Software Studio.

électroniquement dans le registre à instructions qui a ensuite pour tâche de l'*exécuter*. Par exemple, si l'instruction (traduite en anglais) dit « dégager l'accumulateur », l'ordinateur met le nombre 0 dans l'accumulateur ; si l'instruction dit « ajoutez le contenu du registre de mémoire 07 au nombre qui est dans l'accumulateur », l'ordinateur attribue à tout nombre qui figure dans le registre de mémoire l'adresse 07 (les contenus peuvent constituer en n'importe quel nombre) et l'ajoute au nombre qui figure dans l'accumulateur. Et ainsi de suite. Quelles sont les opérations primitives ? Fondamentalement, ce sont les opérations arithmétiques – d'addition, soustraction, multiplication et division –, les opérations de mouvement de données – attribuer, emmagasiner, stocker, entrée, sortie (le cœur de la « logique » des ordinateurs) –, les instructions *conditionnelles*, telles que « SI le nombre qu'il y a dans l'accumulateur est plus grand que zéro, ALORS allez à l'instruction dans le registre 29 ; Sinon, allez à l'instruction suivante. » Selon le modèle de l'ordinateur, il pourrait exister seulement seize opérations primitives ou bien des centaines, toutes câblées dans des circuits à objectifs spécifiques. Chaque opération primitive est codée par une structure binaire unique (par exemple ADDITION pourrait être 1011, et SOUSTRAIRE pourrait être 1101), et chaque fois que ces séquences particulières atterrissent dans le registre à instructions, elles sont comme des numéros de téléphone que l'on compose et qui ouvrent mécaniquement les lignes au circuit spécifique – le circuit d'addition

ou le circuit de soustraction, et ainsi de suite. Les deux registres, dans lesquels une instruction et une valeur seulement peuvent apparaître à n'importe quel moment, forment le célèbre « goulot de von Neumann », le lieu par lequel toute l'activité du système doit passer à la queue leu leu à travers un passage étroit. Sur un ordinateur rapide, il peut exister des millions d'opérations de ce genre en une seconde, et réunies en millions, elles produisent les effets apparemment magiques que peut discerner l'utilisateur.

Tous les ordinateurs digitaux sont les *descendants directs* de cette conception ; et bien que l'on ait introduit de nombreuses modifications et améliorations, ils ont en commun, comme les vertébrés, une architecture sous-jacente fondamentale. Les opérations de base, qui ont l'air d'être arithmétiques, ne semblent pas de prime abord avoir grand-chose à voir avec les « opérations » de base d'un flux de conscience normal – penser à Paris, apprécier l'arôme du pain sorti du four, se demander où l'on ira en vacances. Mais Turing ou von Neumann ne s'en inquiétaient pas. Ce qui comptait à leurs yeux, c'était le fait que cette suite d'actions puisse « en principe » être élaborée de façon à incorporer *toute* « pensée rationnelle » et même peut-être la « pensée irrationnelle ». Il est très ironique, historiquement parlant, que cette architecture ait été décrite de façon trompeuse par la presse populaire à partir du moment où elle a été créée. On a appelé ces fascinantes nouvelles machines de von Neumann des « cerveaux électroniques géants » ; en fait, c'étaient des *esprits électroniques géants*, des imitations électroniques – extrêmement simplifiées – de ce que William James avait appelé le flux de la conscience, cette suite sinueuse de contenus mentaux dont Joyce a donné une description célèbre dans ses romans. L'architecture du cerveau, en revanche, est massivement parallèle ; elle comporte des millions de canaux d'opérations simultanément actifs. Il nous faut comprendre comment un phénomène sériel joycien (ou, comme nous l'avons dit « à la von Neumann ») peut apparaître, avec ses propriétés typiques, dans le brouhaha parallèle du cerveau.

Voici une bien mauvaise idée : nos ancêtres hominiens avaient besoin de penser de façon plus sophistiquée et plus logique ; c'est pourquoi la sélection naturelle a conçu graduellement et installé une machine de von Neumann câblée de façon innée dans l'hémisphère gauche (« logique », « conscient ») du cortex humain. J'espère qu'il est clair, après le récit d'évolution qui vient d'être

donné que bien que ce fait ait pu être *logiquement* possible, il n'a aucune vraisemblance biologique. Nos ancêtres pourraient tout aussi facilement s'être laissé pousser des ailes ou être nés avec des pistolets dans les mains. Ce n'est pas ainsi que l'évolution travaille.

Nous *savons* bien qu'il existe quelque chose qui ressemble ne serait-ce que de loin à une machine de von Neumann dans le cerveau, parce que nous savons que nous avons des esprits conscients « par introspection » ; les esprits que nous découvrons ainsi ont au moins ceci en commun avec les machines de von Neumann : ce sont eux qui ont inspiré les machines de von Neumann ! Ce fait historique a laissé une trace fossile particulièrement frappante : les programmeurs en informatique vous diront qu'il est terriblement difficile de programmer les ordinateurs parallèles qu'on développe aujourd'hui, et relativement facile de programmer une machine sérielle conforme au modèle, de von Neumann. Quand vous programmez une machine de von Neumann classique, vous pouvez utiliser un truc commode : quand cela devient difficile, vous vous demandez ce que vous feriez, si vous étiez la machine, pour résoudre ce problème. Cela vous conduit à répondre : « Bon d'abord, je fais ceci, puis je dois faire cela, etc. » Mais si vous vous demandez ce que vous feriez si vous étiez un processeur parallèle disposant de milliers de canaux, il se produit un blanc ; vous n'avez aucune familiarité avec, ni aucun « accès direct à » des processus qui se produisent dans des milliers de canaux en même temps, même si c'est ce qui se passe en fait dans votre cerveau. Votre seul accès à ce qui se passe dans votre cerveau adopte un « format » séquentiel qui rappelle de façon frappante l'architecture de von Neumann – bien que cette formulation renverse l'ordre historique.

La différence est grande, comme nous l'avons vu, entre une architecture sérielle standard d'ordinateur et l'architecture parallèle du cerveau. On cite souvent ce fait comme une objection à l'intelligence artificielle, qui essaie de créer une intelligence de style humain en construisant des programmes qui marchent (presque toujours) sur des machines de von Neumann. La différence d'architecture est-elle importante ? En un sens non. Turing avait prouvé – et c'est peut-être là sa contribution majeure – que sa Machine de Turing Universelle pouvait calculer n'importe quelle fonction que n'importe quel ordinateur ayant n'importe quelle architecture pourrait calculer. En fait, la Machine

de Turing Universelle est un parfait caméléon mathématique, capable d'imiter *n'importe quelle* autre machine à calculer, capable d'accomplir, durant cette période d'imitation, *exactement* ce qu'effectue cette machine. Pour cela, il suffit de fournir à la Machine de Turing Universelle une description appropriée d'une autre machine ; comme Marcel Marceau (la Machine à Mimer Universelle), dotée d'une chorégraphie explicite, elle se met alors immédiatement à produire une imitation parfaite fondée sur cette description – elle devient, virtuellement, l'autre machine. On peut ainsi considérer un programme d'ordinateur comme une liste d'instructions primitives qu'il faut suivre, ou comme une description de la machine à imiter.

Pouvez-vous imiter Marcel Marceau en train d'imiter un homme ivre lui-même en train d'imiter un joueur de baseball avec sa batte ? Il peut vous sembler que le plus difficile, c'est de garder le contact avec les différents niveaux d'imitation ; mais pour les machines de von Neumann, c'est naturel. À partir du moment où vous avez une machine de von Neumann sur laquelle vous pouvez élaborer de nouvelles machines, vous pouvez enchâsser des machines virtuelles comme des boîtes chinoises. Par exemple, vous pouvez transformer votre machine de von Neumann en machine Unix (le système Unix), puis, implémenter une machine en Lisp (le langage de programmation Lisp) sur la machine Unix – avec Wordstar, Lotus 123, et une quantité d'autres machines virtuelles –, puis implémenter un ordinateur joueur d'échecs sur votre machine en Lisp. Chaque machine virtuelle est reconnaissable par son *interface avec l'utilisateur* – par la façon dont elle apparaît sur l'écran du moniteur et la manière dont elle répond aux entrées. On appelle souvent la façon dont elle se présente *illusion de l'utilisateur*, parce qu'un utilisateur ne peut pas dire – ou ne s'occupe pas de savoir – comment la machine virtuelle particulière qu'il utilise est implémentée dans le matériel. Il lui importe peu que la machine virtuelle soit à un, deux, trois, ou dix niveaux de distance du matériel¹⁷. Par exemple, les utilisateurs de Wordstar peuvent reconnaître, et entrer en interaction avec la machine virtuelle pour Wordstar chaque fois qu'ils la trouvent, quelle que soit la variation qui intervient dans le matériel sous-jacent.

Une machine virtuelle est donc un ensemble temporaire de régularités hautement structurées imposées au matériel sous-jacent par un *programme* : c'est une recette structurée formée

de centaines de milliers d'instructions qui donnent au matériel un énorme ensemble d'habitudes interconnectées et de dispositions à réagir. Si vous considérez les microdétails de toutes les instructions qui passent dans le registre à instructions, vous prendrez les arbres pour la forêt ; mais si vous prenez du recul, vous pourrez voir clairement l'architecture fonctionnelle qui émerge à partir de ces micro-éléments : elle consiste en *entités* virtuelles

telles que des blocs de textes, curseurs, effaceurs, activateurs de peinture, dossiers

et en des *lieux* virtuels

tels que des répertoires, des menus, des écrans, des coquilles reliés par des *chemins virtuels*

tels que ESCAPE vers Dos « ou *entrer dans* le menu » IMPRIMER à *partir* du menu PRINCIPAL

et permettant l'effectuation de vastes et intéressantes *opérations*

telles que chercher un dossier ou un mot, ou élargir une boîte dessinée sur l'écran.

Puisque toute machine à calculer peut être imitée par une machine virtuelle sur une machine de von Neumann, il s'ensuit que si le cerveau est une machine à traitement massivement parallèle, il peut aussi être imité parfaitement par une machine de von Neumann. Dès le début de l'âge des ordinateurs, les théoriciens ont utilisé ce pouvoir caméléonesque des machines de von Neumann pour créer des architectures parallèles *virtuelles* supposées modéliser des structures de type cérébral ¹⁸. Comment transformer une machine qui-fait-une-chose-à-la-fois en machine qui-fait-de-nombreuses-choses-à-la-fois ? Par un processus assez proche du tricot. Supposez que les processeurs parallèles à stimuler aient une largeur de dix canaux. Tout d'abord, on donne à la machine de von Neumann l'instruction d'accomplir les opérations réalisées par le premier nœud du premier canal (le nœud 1 dans le diagramme) en sauvant le résultat dans une mémoire « tampon », puis le nœud 2, et ainsi de suite, jusqu'à ce que les

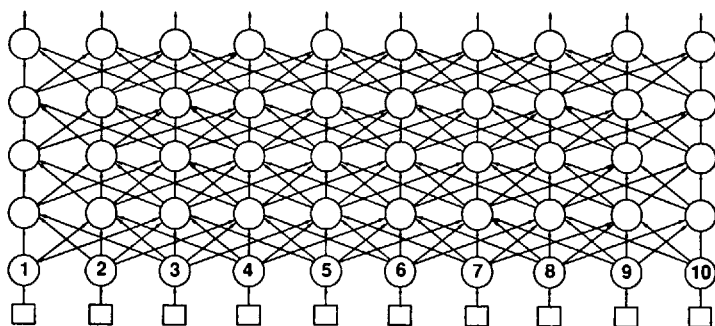


Figure 6.5

nœuds du premier niveau aient avancé d'un moment. Puis, la machine de von Neumann s'attaque aux effets de chacun de ces résultats du premier niveau sur le second niveau de nœuds, en tirant les résultats calculés auparavant, les uns après les autres, de la mémoire tampon, puis les appliquant comme entrées au niveau suivant. Laborieusement, elle travaille, tricotant alternativement, échangeant du temps contre de l'espace. Pour simuler une machine virtuelle à dix canaux, il faut *au moins* dix fois plus de temps que pour une machine à un canal et pour simuler une machine à des millions de canaux (comme le cerveau), il faut au moins un million de fois plus. La démonstration de Turing ne dit rien de la vitesse à laquelle l'imitation peut s'accomplir ; pour certaines architectures, même la vitesse époustouflante des ordinateurs digitaux modernes est dépassée par l'ampleur de la tâche. C'est pourquoi les chercheurs en IA qui s'intéressent aux pouvoirs des architectures parallèles se tournent aujourd'hui vers de *vraies* machines parallèles – artefacts qu'il serait plus juste d'appeler des « cerveaux électroniques géants » – sur lesquelles ils pourraient composer leurs simulations. Mais en principe, n'importe quelle machine parallèle peut être parfaitement mimée – même de façon inefficace – comme une machine virtuelle sur une machine de von Neumann sérielle ¹⁹.

Nous voilà maintenant prêts à renverser cette idée canonique. Vous pouvez simuler un cerveau parallèle sur une machine sérielle de von Neumann ; de même, vous pouvez aussi, en principe, simuler quelque chose comme une machine de von Neumann sur un matériel parallèle. C'est précisément ce que je suggère :

les cerveaux humains conscients sont plus ou moins des machines virtuelles sérielles implémentées – de façon non efficace – sur le matériel parallèle dont l'évolution nous a dotés.

Qu'est-ce qui va jouer le rôle du « programme » quand nous parlons de machine virtuelle tournant sur le matériel parallèle du cerveau ? Il faut une imposante plasticité afin que des myriades de micro-habitudes différentes, et partant de macro-habitudes différentes, soient possibles. Dans le cas de la machine de von Neumann, on utilise des centaines de milliers de zéros et de uns (bits), divisées en « mots » de 8, 16, 32, ou 64 bits, selon la machine utilisée. Les mots sont stockés séparément dans des registres de la mémoire, et on accède à un seul mot à la fois dans le registre à instructions. Dans le cas de la machine parallèle, ce sont, on peut s'en douter, des milliers, des millions ou des billions de dispositifs de connexions renforcées entre les neurones qui seraient utilisés, donnant tous ensemble au matériel sous-jacent un nouvel ensemble de macro-habitudes, un ensemble de régularités conditionnelles de comportement.

Comment ces programmes de millions de connexions renforcées sont-ils installés sur l'ordinateur du cerveau ? Dans une machine de von Neumann, il suffit de « charger » le programme sur la mémoire centrale à partir d'un disque ; l'ordinateur acquiert ainsi instantanément un ensemble de nouvelles habitudes ; dans le cas du cerveau, il faut un *apprentissage*, en particulier l'autostimulation répétitive du type de celle dont nous avons parlé dans la section 5. C'est là, bien entendu, une différence majeure. L'Unité Centrale de Traitement ou CPU (*Central Processing Unit*) d'une machine de von Neumann a une manière rigide de répondre aux suites de bits qui composent ses mots ; elle les traite comme des instructions dans le *langage machine* qui lui est propre et qui est entièrement fixe. Ces faits définissent *l'ordinateur digital à programme incorporé*. Un être humain est totalement différent. Bien qu'il soit probablement vrai que chaque ensemble de connexions renforcées entre neurones dans le cerveau ait un effet déterminé sur le comportement résultant du réseau qui l'entoure, deux cerveaux différents ne peuvent avoir le « même système » d'interconnexions. Dès lors, rien ne peut être comparé au langage machine fixe que partagent, disons, tous les ordinateurs IBM ou compatibles IBM. Par conséquent, si deux cerveaux ou plus « ont un logiciel en commun », ce n'est pas en vertu d'un processus simple,

direct et analogue à celui qui consiste à copier un programme de langage machine d'une mémoire à une autre. (Il est vrai également, bien entendu, que la plasticité qui en quelque sorte sous-tend la mémoire dans un cerveau n'est pas isolée comme un entrepôt passif; la division du travail entre la mémoire et le CPU est un artefact sans équivalent dans le cerveau. Je reviendrai sur ce point au chapitre 8.)

Étant donné qu'il existe de telles différences – qu'on les oublie souvent – pourquoi est-ce que je persiste à rapprocher la conscience humaine d'un logiciel d'ordinateur? Parce que, comme j'espère le montrer, on peut expliquer certaines propriétés cruciales de la conscience qui autrement resteraient problématiques, si l'on forme l'hypothèse que (1) elle est une innovation trop récente pour être câblée au départ dans la machinerie innée, (2) elle est largement un produit de l'évolution culturelle, qu'obtiennent les cerveaux au début de leur apprentissage, et (3) ce sont des myriades de microdispositifs dans la plasticité du cerveau qui déterminent la réussite de son installation, ce qui veut dire que ses traits fonctionnellement importants sont très vraisemblablement invisibles à l'analyse neuro-anatomique en dépit de la saillance extrême des effets. Aucun informaticien ne tenterait de comprendre les différentes forces et faiblesses de Wordstar comparé à Wordperfect en s'appuyant sur les différences dans les structures de voltages de la mémoire. De même, aucun chercheur en sciences cognitives ne devrait s'attendre à donner un sens à la conscience humaine simplement en s'appuyant sur la neuro-anatomie. De plus (4), l'idée qu'il existe une *illusion de l'utilisateur* d'une machine virtuelle est terriblement suggestive : si la conscience est une machine virtuelle, qui est l'utilisateur qui a cette illusion visuelle? Je gage que cela ressemble étrangement à un retour inexorable à un Moi cartésien assis dans la gare de triage corticale et réagissant à l'illusion de l'utilisateur du logiciel qui y fonctionne. Mais on peut, comme nous allons le voir, échapper à ce dénouement effrayant.

Supposez donc pour le moment qu'on dispose d'une version plus ou moins bien conçue (sans bogues) de cette machine virtuelle que représente le flux de la conscience – la machine joycienne – dans ma mémosphère. Comme nous l'avons vu, étant donné qu'aucun langage machine n'est commun aux différents cerveaux, les méthodes de transmission qui garantiraient l'opération assez uniforme d'une machine virtuelle à travers la culture

doivent être des méthodes sociales, très contextualisées, et à un certain degré pourvues d'une auto-organisation et d'une auto-correction. Faire « parler ensemble » deux ordinateurs différents – par exemple un Macintosh et un IBM-PC – est un problème d'ingénierie délicat et complexe, qui dépend de l'information précise dont on dispose sur la machinerie interne des deux systèmes. Dans la mesure où les êtres humains peuvent « avoir un logiciel en commun » sans que qui que ce soit le sache, ce doit être parce que les systèmes communs ont un haut degré de labilité et de tolérance des divers formats. Il existe plusieurs méthodes pour mettre en commun ces logiciels : l'apprentissage par imitation, l'apprentissage par « renforcement » (soit imposé délibérément par le maître – récompense, encouragement, désapprobation, menace – ou subtilement et inconsciemment transmis à travers les rencontres communicatives) et l'apprentissage comme résultat d'une instruction explicite formulé dans une langue naturelle qui ait déjà été apprise par les deux premières méthodes. (Pensez, par exemple, aux types d'habitudes qui se formeraient en disant souvent, à un novice, « Dis-moi ce que tu es en train de faire » et « Dis-moi pourquoi tu fais cela ». Il suffit alors d'imaginer que le novice prenne l'habitude de se poser les mêmes questions à lui-même.)

En fait, je soupçonne que le langage parlé mais aussi l'écriture jouent un rôle majeur dans le développement et l'élaboration des machines virtuelles que la plupart d'entre nous faisons marcher dans nos cerveaux. La roue est une belle technologie dont l'utilité dépend beaucoup des rails, des chaussées pavées ou d'autres surfaces aplanies artificiellement ; de même, la machine virtuelle dont je suis en train de parler ne peut exister que dans un environnement qui n'a pas seulement le langage et l'interaction sociale, mais aussi l'écriture et les diagrammes. Les exigences de la mémoire et de la reconnaissance des structures qui conditionnent son implantation requièrent en effet que le cerveau « décharge » certains de ses souvenirs dans des tampons qui se trouvent dans l'environnement. (Notez que cela implique que la « mentalité d'avant l'écriture » pourrait bien avoir impliqué une classe d'architectures virtuelles passablement différentes de celles que l'on rencontre dans des sociétés qui connaissent l'écriture.)

Réfléchissez à ce qui se passe quand vous additionnez mentalement deux nombres à dix chiffres sans utiliser ni papier ni stylo et sans prononcer les nombres à haute voix. Imaginez-vous

en train de vous représenter, sans l'usage d'un diagramme, une façon de faire se rencontrer trois branches d'autoroutes par une intersection en forme de feuille de trèfle en sorte que l'on puisse aller à partir de n'importe quelle direction sur l'une des branches d'autoroute vers n'importe quelle autre branche sans avoir à entrer sur la troisième branche. Ce sont des problèmes que les êtres humains résolvent aisément avec l'aide de dispositifs de mémoire externe et grâce à leurs scanners innés (appelés *yeux* et *oreilles*) à leurs circuits innés et très développés de reconnaissance des formes. (Voir Rumelhart, chapitre 14, dans Mc Lelland et Rumelhart, 1986 pour quelques bonnes observations sur ce sujet.)

Nous installons dans nos cerveaux, au moment de notre développement pendant la prime enfance, un ensemble d'habitudes d'esprit organisées et partiellement pré-testées (comme dit le théoricien des sciences politiques Howard Margolis (1987)). Nous considérerons plus en détail cette architecture au chapitre 8. Pour l'instant, je voudrais suggérer que la structure globale du nouvel ensemble de régularités est une *mise en chaîne* sérielle, dans laquelle d'abord une première « chose », puis une autre « chose » a lieu (en gros) au même « endroit ». Ce flux d'événements est entraîné par une quantité d'habitudes acquises ; celle de se parler à soi-même en est le principal exemple.

Étant donné que la nouvelle machine qui est créée en nous est un complexe de mêmes hautement répliqués, nous pouvons nous demander à quoi il doit son succès répliatif. Nous devons bien entendu garder à l'esprit qu'il *pourrait* ne servir à rien, sinon à faire des répliations. Cela *pourrait* être un virus dans le logiciel, qui parasite facilement les cerveaux humains sans donner effectivement aux êtres humains dont il infeste les cerveaux un avantage quelconque dans la compétition. Ou, ce qui est plus plausible, *certaines traits* de la machine pourraient être des parasites, qui n'existent que parce qu'ils peuvent s'en débarrasser, et parce que ce n'est pas possible – ou parce que cela n'en vaut pas la peine. William James pensait qu'il est absurde de supposer que la chose la plus étonnante que nous connaissions dans l'univers – la conscience – soit un simple artefact, qui ne joue aucun rôle essentiel dans la manière dont notre cerveau fonctionne. Mais, aussi improbable que cela puisse paraître, ce n'est pas totalement hors de question. Par conséquent, ce n'est pas totalement absurde. Nous avons autour de nous quantité de

preuves des bénéfices que la conscience semble nous apporter. C'est pourquoi, sans aucun doute, nous pouvons nous satisfaire de ses diverses *raisons d'être*. Mais nous interprétons ces preuves de travers si nous pensons que le mystère subsistera à moins que chaque propriété particulière ne reçoive – ou n'ait reçu dans le passé – une fonction (de notre point de vue en tant qu'« utilisateurs » de la conscience) (Harnad, 1982). Il peut très bien se faire que certains faits bruts ne soient pas justifiés. Certains traits de la conscience pourraient bien être des mêmes égoïstes. Mais si nous envisageons le bon côté des choses, quels sont les problèmes que cette nouvelle machine est supposée résoudre ? Le psychologue Julian Jaynes (1976) nous a persuadés que ses capacités d'auto-exhortation et d'automémorisation étaient des conditions nécessaires pour qu'aient pu s'organiser les types d'autocontrôle à long terme sans lesquels l'agriculture, le fait de faire des projets, et autres activités civilisées et civilisatrices seraient impossibles. Ces traits semblent aussi requis pour le type d'autogestion qui peut protéger un système déficient face à ses propres échecs – c'est un thème qu'a développé en intelligence artificielle Douglas Hofstadter (1985). Et le psychologue Nicholas Humphrey (1976, 1983a, 1986) a considéré ce thème comme fournissant un moyen d'exploiter ce que l'on pourrait appeler des stimulations sociales – utiliser l'introspection pour guider nos intuitions quant à ce que les autres pensent et ressentent.

Derrière ces talents plus spécialisés et plus avancés, se trouve la capacité de résoudre les métaproblèmes qui se posent au sujet de ce qu'il faut penser ensuite. Nous avons vu plus haut dans ce chapitre que quand un organisme fait face à une crise (ou simplement un problème nouveau et délicat), il peut contenir des ressources qui seraient très utiles en ces circonstances *si seulement il pouvait les trouver et les utiliser au bon moment* ! Les réponses d'orientation, comme l'a suggéré Odmar Neumann, ont pour effet bénéfique de faire venir presque tout le monde en même temps, mais le fait d'accomplir cette levée globale, comme nous l'avons vu, fait autant partie du problème que de la solution. Cela n'est d'aucune utilité tant qu'à l'étape suivante le cerveau ne parvient pas à obtenir une activité cohérente quelconque de la part de tous ces volontaires. Le problème dont les réponses d'orientation constituaient une solution partielle était celui-ci : obtenir un accès total, global dans une collection de spécialistes habitués à s'occuper de leur propre boulot. Même si, grâce à une

architecture sous-jacente de type Pandémonium, le chaos se stabilise bientôt, laissant temporairement les commandes à un seul spécialiste (qui soit peut-être mieux informé quant à la compétition qu'il a engagée), il existe, de toute évidence, au moins autant de façons dont ces conflits peuvent se résoudre mal qu'il y a de manières dont ils peuvent se résoudre bien. Rien ne garantit que le spécialiste le plus politiquement efficace soit l'« homme de la situation ».

Platon a compris ce problème il y a deux mille ans, et il a trouvé une métaphore merveilleuse pour l'exprimer :

« Vois dès lors si, pour la connaissance aussi, il se peut de la sorte qu'on la possède sans l'avoir, et si tout ne se passe pas comme si, ayant pris à la chasse des colombes sauvages, ou tel autre oiseau, on les élevait chez soi dans un colombier qu'on aurait installé. En un sens, nous dirions en effet alors, je pense, qu'on a toujours ses colombes, justement parce qu'on les possède : n'est-il pas vrai ?... Mais en un autre sens, nous dirions bien qu'on n'en a pas une seule, mais plutôt que, du fait qu'on se les ait mises à portée de la main dans une enceinte appropriée, une possibilité nous est survenue à leur égard, possibilité de les attraper et de les tenir quand on le voudra, après qu'on se sera mis en chasse de celle qu'à chaque fois on désire, puis, au rebours, de la lâcher. Et c'est ce qu'il nous est possible de faire autant de fois que bon nous semblera... Voilà donc qu'il nous faut construire [dans l'intérieur de l'âme] un colombier contenant des oiseaux de diverses sortes : ceux-ci sont par groupes, isolés les uns des autres, ceux-là forment des groupes restreints ; quelques-uns enfin voltigent tout seuls, comme cela se trouve, à travers tous les autres » (*Théétète*, 197-198a, trad. Robin ²⁰).

Platon a vu que le fait de disposer seulement d'oiseaux ne suffit pas ; la difficulté est d'apprendre comment attraper le bon oiseau quand vous l'appellez. Selon lui, en *raisonnant*, nous améliorons notre capacité à obtenir la venue des bons oiseaux au bon moment. Apprendre à raisonner, c'est en fait apprendre des stratégies de rétention des connaissances ²¹. C'est là qu'interviennent les habitudes d'esprit. Nous avons déjà vu en gros comment de telles habitudes d'esprit générales telles que se parler à soi-même ou se faire des diagrammes à soi-même *pourraient* faire venir les bonnes séquences d'informations à la surface. (À la surface de quoi ? C'est un sujet que je remets au chapitre 9.) Mais adopter des habitudes d'esprit plus spécifiques, raffiner et

améliorer la façon de se parler à soi-même peut augmenter encore plus les chances de quelqu'un.

Le philosophe Gilbert Ryle, dans son livre posthume intitulé *On Thinking* (1979), a affirmé que la pensée, la pensée lente, difficile, pondérée, du type de celle que la fameuse statue de Rodin est apparemment en train d'accomplir, doit en fait souvent être une sorte de discours adressé à soi-même. Surprise, surprise ! N'est-il pas évident que c'est ce que nous semblons (souvent) faire ? Nous pouvons même souvent nous communiquer les différents mots que nous exprimons dans nos soliloques silencieux. Mais ce qui est loin d'être évident, c'est la raison pour laquelle se parler à soi-même est utile.

« Qu'est-ce que *Le Penseur* est en train de faire, et qui se passe apparemment dans son intériorité cartésienne ? Ou, pour faire scientifique, à quoi ressemblent les processus mentaux qui se produisent dans cette *camera obscura* cartésienne... Il est notoire que certaines de nos cogitations, mais pas toutes, trouvent leur aboutissement dans la solution de nos problèmes ; nous étions dans le brouillard mais finalement nous sommes au clair. Mais si nous réussissons quelquefois, pourquoi cela ne se produit-il pas toujours ? Si nous y parvenons en dernière instance, pourquoi n'y parvenons-nous pas rapidement ? Si c'est avec difficulté, pourquoi n'est-ce pas avec facilité ? Pourquoi est-ce que cela marche dans certains cas ? Et comment cela peut-il marcher ? » (Ryle, 1979, p. 65).

Les habitudes d'esprit se sont forgées à travers les siècles pour former les passages à travers des sentiers d'exploration battus. Comme le note Margolis,

« Même un être humain aujourd'hui (et par conséquent, *a fortiori*, un ancêtre lointain des êtres humains contemporains) ne peut pas aisément ni même ordinairement maintenir son attention constante sur un problème unique pendant plus de quelques dizaines de secondes. Et pourtant nous travaillons sur des problèmes qui requièrent bien plus de temps. La manière dont nous le faisons (comme nous pouvons l'observer en nous regardant nous-mêmes) requiert des périodes de rumination suivies de périodes de récapitulation, dans lesquelles nous nous décrivons à nous-mêmes ce qui semble s'être produit durant la rumination, qui conduisent à tel ou tel résultat. La fonction de ce phénomène est évidente : en reprenant ces résultats intermédiaires..., nous les impliquons dans la mémoire, parce que les contenus immédiats du flux de la conscience sont très vite perdus si nous ne les reproduisons pas... Avec

le langage, nous pouvons décrire à nous-mêmes ce qui a semblé se produire pendant la rumination qui a conduit à un jugement, produire une version reproductible du processus d'atteinte d'un jugement, et l'impliquer dans la mémoire à long terme en le reproduisant en fait » (Margolis, 1987, p. 60).

C'est là, dans les habitudes individuelles d'autostimulation, que l'on doit rechercher des *kludges*, le terme pirate en informatique pour désigner les processus *ad hoc* de dépistage qu'on met en général dans le logiciel pour enlever les bogues et faire marcher l'ensemble du processus. (La linguiste Barbara Partee a un jour dit d'un dispositif élégant dans un programme d'IA analyseur de langage que c'était un « *odd hack* » (*un drôle de piratage*.) C'est la plus jolie contrepèterie que j'aie jamais rencontrée. Dame Nature est pleine de piratages bizarres, et nous devons nous attendre à les trouver aussi dans l'adoption idiosyncrasique de la machine virtuelle par un individu.)

Voici un exemple plausible. Puisque la mémoire humaine n'est pas conçue de façon innée pour être une mémoire totalement fiable, d'accès rapide et aléatoire (ce que requiert toute machine de von Neumann), quand les concepteurs (culturellement et temporellement distribués) de la machine virtuelle à la von Neumann se sont trouvés confrontés à la tâche de fabriquer un substitut approprié qui puisse marcher sur un cerveau, ils sont tombés sur divers Trucs permettant d'améliorer la mémoire. Les Trucs de base sont la reproduction, la reproduction, et encore la reproduction, encouragée par des rimes et des rythmiques, des maximes qui soient faciles à faire revenir en mémoire. (Les rimes et les rythmes exploitent le vaste pouvoir du système pré-existant d'analyse auditive pour reconnaître des trames de sons.) La juxtaposition répétée délibérée des éléments au sein desquels on avait besoin de construire un lien d'association – de manière à ce qu'un élément « rappelle » toujours au cerveau celui qui suit – s'est trouvée encore améliorée, nous pouvons le supposer, en rendant les associations aussi riches que possible, en les habillant non seulement de caractéristiques visuelles et auditives, mais en exploitant le corps tout entier. Le froncement de sourcils du *Penseur* et les poses que nous pouvons favoriser à titre d'idiosyncrasies – se tenir le menton, se gratter la tête, murmurer, marcher de long en large et griffonner sur des bouts de papier – pourraient se révéler n'être pas seulement des sous-produits accidentels de la pensée consciente, mais des éléments fonction-

nels (ou des traces vestiges de contributions fonctionnelles antérieures) contribuant à la soumission laborieuse du cerveau à une discipline nécessaire à sa transformation en un esprit venu à maturité.

À la place des cycles précis et systématiques de recherche ou d'instruction qui amènent chaque instruction nouvelle au registre à instructions en vue de leur exécution, nous devons chercher des « règles » de transition mal ficelées, un peu hasardées, et non logiques, où le penchant largement inné du cerveau pour l'« association libre » reçoit des chaînes d'association longues qui assurent plus ou moins que les bonnes séquences ont été essayées. (Au chapitre 8, nous considérerons des développements de cette idée en 1A ; pour des analyses complémentaires insistant sur tel ou tel point, voir Margolis, 1987, et Calvin, 1987, 1989. Voir aussi Dennett 1991b.) Nous ne devons pas nous attendre à ce que la plupart des séquences qui se sont produites soient des *algorithmes* bien établis, supposés garantir les résultats recherchés, mais simplement des sondages approximatifs dans le colombier de Platon.

L'analogie entre la conscience et les machines virtuelles de l'informatique nous fournit une perspective éclairante sur ce phénomène. On a supposé au départ que les ordinateurs n'étaient que des machines à broyer des nombres, mais aujourd'hui leur broyage de nombres a été transformé en une multitude de méthodes ingénieuses pour créer de nouvelles machines virtuelles, telles que les jeux vidéo et les traitements de texte, dans lesquels le broyage des nombres est presque invisible et qui semblent posséder des pouvoirs nouveaux et quasiment magiques. De même, nos cerveaux n'étaient pas faits (sauf pour certains organes périphériques récents) pour s'adonner au traitement de texte ; mais à présent une large portion – peut-être même la part du lion – de l'activité qui se produit dans les cerveaux humains adultes est impliquée dans une sorte de traitement de texte : production du langage et compréhension, reproduction sérielle et réarrangement d'entités linguistiques, ou mieux encore, de leurs substituts neuronaux. Ces activités magnifient et transforment les pouvoirs du matériel sous-jacent comme par magie (si on regarde les choses de l'extérieur).

Mais pourtant (je suis sûr que vous avez envie de faire cette objection) : tout cela n'a que peu à voir ou n'a même rien à voir avec la conscience ! Après tout, une machine de von Neumann

est entièrement inconsciente ; pourquoi le fait de l'implémenter – ou d'implémenter une machine joycienne – serait-il plus conscient ? J'ai en fait une réponse : la machine de von Neumann, du fait qu'elle est câblée dès le départ de cette façon, avec des liens informationnels maximalement efficaces, n'a pas eu besoin de devenir l'objet de ses propres systèmes perceptuels élaborés. Les mécanismes de la machine joycienne, d'un autre côté, sont tout aussi « visibles » et « audibles » pour celle-ci que n'importe quelle chose du monde extérieur qu'elle est conçue pour percevoir – pour la simple raison qu'elles ont dans une large mesure la même machinerie perceptuelle tournée vers eux.

Cela *semble* être un jeu de miroirs, je sais. Et cela choque sans aucun doute nos intuitions. C'est difficile à avaler, et au départ, cela nous rebute. C'est précisément ce à quoi l'on peut s'attendre de la part d'une idée qui pourrait briser des siècles de mystère, de controverse et de confusion. Dans les deux chapitres qui viennent, nous examinerons plus précisément – sans nous départir de notre scepticisme – comment ce qui a l'air d'être un truc composé de miroirs peut expliquer en partie la conscience.

NOTES

1. Vous pouvez inférer que tout ce que j'utilise dans mon récit est correct – ou tout du moins sur la bonne voie – mais vous ne devez pas inférer que du fait que je laisse de côté une théorie quelconque ou certains aspects d'une théorie, que je crois que cette théorie est fausse. Vous ne devez pas inférer non plus que simplement parce que j'utilise quelques détails d'une théorie, je crois que le reste de la théorie est défendable. Cela s'applique aussi à mes propres écrits sur ces sujets, auxquels j'emprunte un certain nombre de points.

2. John Maynard Smith est le théoricien contemporain le plus en vue et, outre son livre classique *The Evolution of Sex* (1978), on trouve plusieurs articles brillants sur les problèmes conceptuels dans son recueil d'articles, *Sex, Games and Evolution* (1989). Voir aussi R. Dawkins (1976, p. 46-48), pour un bref survol de ces questions.

3. L'idée de neurones multifonctionnels n'est pas nouvelle, mais elle a gagné récemment des adhérents :

Ce sont les concaténations plus ou moins simultanées de sorties neuronales ou de signaux qui sont ambiguës, plutôt que les sorties de neurones individuels. La convergence de diverses concaténations de signaux ambigus à chaque niveau successif résoudrait en partie l'ambiguïté tout comme la convergence des définitions ambiguës détermine une solution presque unique à des problèmes de mots croisés (Dennett, 1969, p. 56).

Il n'y a pas de structure ni de combinaison unique de groupes correspondant à une catégorie unique ou trame d'output. Au lieu de cela, plus d'une combinaison de groupes neuronaux peut produire un output particulier, et un groupe simple

donné peut participer à plus d'un type de fonction de signalisation. Cette propriété des groupes neuronaux dans des répertoires, appelée *dégénérescence*, fournit l'une des bases fondamentales pour la capacité généralisante des cartes de réentrée (Edelman, 1989, p. 50).

Ce trait architectural, dans lequel chaque nœud contribue à de nombreux contenus distincts, était déjà souligné par Hebb dans son ouvrage pionnier *The Organization of Behavior : A Neuropsychological theory* (1949). Il est au cœur du « traitement parallèle distribué » ou du « connexionnisme ». Mais la notion de multi-fonction tient à autre chose : à un niveau plus grossier d'analyse, nous verrons des systèmes tout entiers avoir des rôles spécialisés, mais aussi être susceptibles d'être recrutés dans des projets plus généraux.

4. L'analogie entre l'ascidie et le maître de conférences fut proposée, à ma connaissance, par le neurobiologiste Rodolfo Llinas.

5. Ce type de câblage rappelle la façon pas tout à fait fiable dont Shakey distinguait les boîtes des pyramides. Par conséquent, Shakey n'est finalement pas totalement un être dénué de traits biologiques ; la biosphère a quantité de gadgets de ce genre. Il reste vrai, cependant, que le système « visuel » de Shakey n'est pas du tout un bon modèle de la vision d'une espèce vivante quelconque. Il n'était pas fait pour cela.

6. On peut discerner les idées de base dans les écrits de Darwin et de ses premiers disciples (Richards, 1987). Le neuro-anatomiste J.Z. Young (1965a, 1965b) fut le pionnier d'une théorie sélectionniste de la mémoire (voir aussi Young, 1979). J'ai développé une version philosophique de l'argument de base, en esquissant certains détails, dans ma thèse à Oxford, 1965, dont une version abrégée, « Evolution in the brain », est devenue le chapitre 3 de *Content and Consciousness*, 1969. John Holland (1975) et d'autres chercheurs en intelligence artificielle, ont développé des « algorithmes génétiques » pour la reconstruction de systèmes d'apprentissage (voir aussi Holland, Holyoak, Nisbet, Thagard, 1986) et Jean-Pierre Changeux (Changeux et Danchin, 1976 ; Changeux et Dehaene, 1989) ont proposé un modèle neuronal assez détaillé. Le neurobiologiste William Calvin (1987, 1989a) fournit une perspective assez différente (et plus aisément accessible) sur ces questions avec sa propre théorie de l'évolution dans le cerveau. Voir aussi son compte rendu clair et lucide (Calvin 1989b) du livre de Gerald Edelman, *Neural Darwinism* (1987). Plus récemment, Edelman a publié *the Remembered Present : A biological Theory of Consciousness* (1989).

7. Trad. Caillols (NdT).

8. C'est le problème fondamental, en philosophie de l'esprit, du *contenu mental* ou de l'*intentionnalité*, et les solutions qu'on a proposées à ce problème sont, c'est bien connu, sujettes à controverse. J'ai fourni la mienne dans *La Stratégie de l'interprète*.

9. Quelques théoriciens intrépides ont soutenu que c'est faux. Jerry Fodor (1975), par exemple, a soutenu que tous les concepts que nous puissions avoir sont donnés de façon innée, et qu'ils doivent seulement être déclenchés ou rendus accessibles par des expériences particulières d'« apprentissage ». Ainsi Aristote avait-il le concept d'un avion dans son cerveau, et aussi celui d'une bicyclette – il se trouve seulement qu'il n'avait jamais eu l'occasion de les utiliser ! À ceux qui éclatent de rire entendant formuler cette idée, Fodor répond que les immunologues se gaussaient de l'idée que les gens – Aristote, par exemple – sont nés avec des millions d'anticorps différents, y compris des anticorps spécifiques à des composés qui ne sont apparus dans la nature qu'au XX^e siècle, mais ils ne se gaussent plus aujourd'hui : ce fait est établi. Le problème, avec cette théorie, quand on l'applique à la fois à l'immunologie et à la psychologie, est que ses versions radicales sont de toute évidence fausses et que ses versions plus modérées sont indistinguables de la thèse opposée. Il y a bien une réaction combinatoire dans le système immunitaire – toute réponse innée n'est pas une correspondance bi-univoque entre des types simples d'anticorps préexistants ; de manière ana-

logue, il se peut qu'Aristote ait eu un concept inné d'*avion*, mais avait-il aussi un concept inné de *gros jumbo-jet* ? Avait-il le concept d'un *aller et retour en APEX de Boston à Londres* ? Quand on distingue ces questions, dans les deux cas, il se trouve qu'il y a quelque chose comme un apprentissage dans les deux cas, et quelque chose comme des concepts innés dans les deux cas.

10. Je fais allusion, bien entendu, à la théorie de la signification non naturelle de Paul Grice (Grice, 1957, 1969), mais pour une nouvelle théorie de la communication qui remplace certains des traits les plus fragiles et les plus improbables des théories de Grice, voir Sperber et Wilson (1986).

11. De quel droit puis-je parler de croyances et de désirs de ces ancêtres non-encore-complètement-conscients ? Ma propre théorie du désir et de la croyance, exposée dans *La Stratégie de l'interprète*, défend la thèse selon laquelle il n'y a pas de raison valable de mettre ces termes entre guillemets : le comportement d'animaux « inférieurs » (même les grenouilles) est un domaine auquel s'applique tout autant l'explication par la posture intentionnelle, avec son imputation de désirs et de volontés, que le comportement des êtres humains. Mais les lecteurs qui ne sont pas d'accord avec cette théorie peuvent comprendre les termes qui vont être utilisés ici en un sens métaphorique étendu.

12. Sur la communication chez les primates et sur la question toujours non résolue de savoir si oui ou non les singes et les grands singes sont capables de tromperie délibérée (voir Dennett, 1983, 1988c, 1988d, 1989a ; Byrne and Whiten, 1988 ; Whiten and Byrne 1988).

13. Dans « Le jardin aux sentiers qui bifurquent », Jorge Luis Borges (1962) invente une version diaboliquement habile de cette stratégie, que je m'abstiendrai de décrire, ne voulant pas vendre la mèche sur la fin de l'histoire.

14. J'emprunte cette section à mon article « Les mêmes et l'exploration de l'imagination » (1990a).

15. Pour un bon exposé récent des controverses à ce sujet dans la littérature spéculant sur l'évolution du langage, voir Pinker et Bloom (1990), et les commentaires afférents.

16. Les puristes pourraient objecter que mon usage du terme *machine virtuelle* est quelque peu plus large que l'usage qu'ils recommandent en informatique. Je réponds que, comme Dame Nature, quand je vois une entité disponible que je peux « prélever » pour en faire un usage étendu (Gould, 1980), je m'en sers.

17. Ou bien cela pourrait ne pas être une machine virtuelle du tout. Cela pourrait être une machine réelle câblée pour un but spécifique pour répondre aux ordres, comme une machine en Lisp, qui est un descendant des machines *virtuelles* en Lisp, et qui est conçue, jusqu'à ses puces de silicone pour programmer le langage Lisp.

18. Les « neurones logiques » de Mc Culloch et Pitts (1943) furent en fait conçus en même temps que l'intention de l'ordinateur sériel, et ils influencèrent la pensée de von Neumann, et tout cela en retour conduisit aux Perceptrons des années cinquante, les ancêtres du connexionnisme d'aujourd'hui. Pour une brève analyse historique, voir Papert (1988).

19. Pour en savoir plus sur les implications de la vitesse réelle, et pour ses implications pour l'intelligence artificielle, voir « Pensée rapide » dans mon livre *La Stratégie de l'interprète* (1987a).

20. *Minus* un membre de phrase omis par Dennett (NdT).

21. Pour une discussion intéressante du désaccord (apparent) entre deux écoles de pensée en intelligence artificielle, le raisonnement versus la recherche, voir Simon et Kaplan, 1989, p. 18-19.

Chapitre 7

Ce que les dires peuvent nous faire

« Le langage, comme la conscience, ne surgit que du besoin, de la nécessité, de l'échange avec les autres. »

Karl Marx, 1846.

« La conscience généralement ne s'est développée que sous la pression de la nécessité de communiquer. »

Frédéric Nietzsche, 1882.

« Avant que mon maître ne vienne, je ne savais pas que j'existais. Je vivais dans un monde qui était un non-monde. Je ne peux pas espérer décrire adéquatement ce temps inconscient, et pourtant conscient de néant [...] Comme je n'avais pas de pouvoir de penser, je ne comparais pas un état mental avec un autre. »

Helen Keller, 1908.

Faisons le point : e pluribus unum ?

Au chapitre 4, nous avons exposé l'idée aussi séduisante que mauvaise du Théâtre Cartésien, dans lequel on présente un spectacle en son et lumière à un public solitaire mais puissant, l'Ego ou l'Administration centrale. Même si nous avons vu combien cette idée était incohérente, et envisagé une théorie rivale, le modèle des Versions Multiples, le Théâtre Cartésien continuera

à nous hanter tant que nous n'aurons pas ancré fermement notre théorie rivale au sol ferme des sciences empiriques. Cette entreprise a commencé au chapitre 5, et au chapitre 6 nous avons encore progressé. Nous sommes revenus, littéralement, aux premiers principes : ceux de l'évolution qui nous ont dirigé dans notre récit spéculatif du processus graduel de développement de l'architecture qui est à l'origine du type de conscience qui est la nôtre. Essayons de jeter un coup d'œil sur la machinerie de la conscience qui se trouve à l'intérieur de la boîte noire. Partons des coulisses, pour employer une expression rendant hommage à la métaphore théâtrale dont nous essayons de nous débarrasser.

Il existe dans notre cerveau un assemblage de circuits cérébraux rafistolés qui, grâce à une famille d'habitudes variées que nous ont inculqués pour une part la culture et pour une autre part nos explorations individuelles, conspirent à créer une machine virtuelle plus ou moins réglée, plus ou moins efficace et plus ou moins bien construite, la *machine joycienne*. En rassemblant ces organes qui ont évolué indépendamment de manière à leur faire servir une cause commune et en dotant leur réunion de pouvoirs énormes, cette machine virtuelle, ce logiciel du cerveau, accomplit une sorte de miracle politique interne : il crée un *capitaine virtuel* de l'équipage, sans donner à aucun de ces organes un pouvoir dictatorial à long terme. Qui commande ? D'abord une coalition, puis une autre ; elles se passent le relais de manière non chaotique par l'effet de bonnes vieilles méta-habitudes produisant des suites cohérentes et sensées plutôt qu'une mêlée interminable pour la course au pouvoir.

Le sage commandement qui en résulte n'est que l'un des pouvoirs qu'on attribue traditionnellement au Moi, mais il n'en est pas moins important. William James lui a rendu hommage quand il a brocardé l'idée d'un Neurone Pontifical qui se trouverait quelque part dans le cerveau. Nous savons que la petite annonce qui décrit l'emploi du sous-système qui sera le Chef dans le cerveau est incohérente, mais nous savons aussi que ces responsabilités et ces décisions doivent être empaquetées *d'une manière ou d'une autre* dans le cerveau. Nous ne sommes point comme des vaisseaux allant au hasard et menés par des équipages bagarreurs ; nous ne nous en tirons pas en nous gardant seulement des hauts-fonds et d'autres dangers ; nous menons des campagnes planifiées, nous corrigeons des erreurs tactiques, nous reconnaissons les messages subtils qui annoncent les possibilités

qui nous sont ouvertes, et nous contrôlons d'énormes projets qui mettent des mois à s'accomplir. Dans les chapitres qui vont suivre, nous allons essayer d'examiner de plus près l'architecture de cette machine virtuelle, pour essayer d'étayer, à défaut de prouver, notre hypothèse selon laquelle cette machine pouvait effectivement avoir ces fonctions directrices ainsi que d'autres. Mais avant de nous atteler à cette tâche, il nous faut exposer et neutraliser une autre source possible de mystification : l'illusion du Signifieur Central.

L'une des tâches centrales du Chef imaginaire consiste à contrôler la communication avec le monde extérieur. Comme nous l'avons vu au chapitre 3, l'idéalisation qui rend possible l'hétérophénoménologie *présuppose* qu'il existe quelqu'un à la maison qui soit là pour parler, un Auteur de l'enregistrement, un Signifieur de toutes les significations. Quand nous entreprenons d'interpréter les sons vocaux émis par un corps, nous ne supposons pas que ce sont seulement des jappements désordonnés, ou des mots tirés d'un chapeau par un troupeau de fêtards en coulisses. Ils nous semblent être les actes d'un seul agent, la (seule et unique) personne dont le corps produit ces sons. Si nous choisissons vraiment d'interpréter, nous n'avons pas d'autre choix que de postuler une personne dont nous interprétons les actes de communication. Cela ne revient pas tout à fait à postuler l'existence d'un système *interne* qui serait le chef du corps, le Maître de Marionnettes contrôlant la marionnette, mais c'est l'image qui nous vient naturellement à l'esprit. Le Chef interne, peut-on supposer, est plutôt comme le président des États-Unis : il peut demander à un porte-parole ou à d'autres subordonnés de produire des communiqués de presse, mais quand ils parlent, ceux-ci parlent en son nom, exécutent les actes de langage dont il est responsable et dont il est officiellement l'auteur.

En fait, aucune chaîne de commandement dans le cerveau ne gouverne la production du langage (ou de l'écriture, en l'occurrence). Si nous voulons démonter le Théâtre Cartésien, nous devons en particulier expliquer de manière plus plausible les sources effectives des assertions, des questions et d'autres actes de langage que nous attribuons naturellement à la personne (unique) dont le corps émet les énonciations. Nous devons savoir comment se comporte notre mythe heuristique de l'hétérophénoménologie quand nous prenons en compte les complexités liées à la production du langage.

Nous avons déjà pu apercevoir l'ombre portée de ce problème. Au chapitre 3, nous avons imaginé que le robot Shakey avait une capacité rudimentaire à faire la conversation ou tout du moins à émettre des mots en diverses circonstances. On pourrait faire en sorte que Shakey « nous dise » comment il différencie les boîtes des pyramides. Shakey pourrait dire : « Je passe en revue toutes les suites de dix mille chiffres... » ou « Je trouve les frontières entre clair et sombre et trace une ligne... » ou « Je ne sais pas ; certaines choses ont simplement l'air d'avoir la forme de boîtes... ». Chacun de ces différents « comptes rendus » émis à partir de différents niveaux d'accès que pourrait avoir la machinerie à faire des « comptes rendus » pourrait avoir à aller en direction de la machinerie destinée à identifier les boîtes, mais nous ne sommes pas entrés dans les détails des manières dont les différents états de la machine pourraient être liés aux productions imprimées qu'elles ont émises. Ce modèle délibérément simplifié de la production du langage n'avait d'autre utilité que de nous permettre d'avancer une idée très abstraite sur la base d'une expérience de pensée : si un système chargé d'émettre des phrases n'avait qu'un accès limité à ses états internes et un vocabulaire limité avec lequel composer ses phrases, on ne pourrait interpréter ses « comptes rendus » comme vrais que si on l'analysait en un sens assez métaphorique. Les « images » de Shakey nous ont fourni un exemple de la manière dont quelque chose qui n'était pas en fait une image pourrait être la chose même à propos de laquelle on parle en fait, sous la forme d'une image.

Envisager une possibilité abstraite est une chose ; montrer que cette possibilité peut s'appliquer à nous de façon plausible en est une autre. Shakey ne faisait pas vraiment des comptes rendus, il ne faisait pas de vraies énonciations. Pour autant que nous puissions nous en rendre compte, la verbalisation que nous imaginions chez Shakey pourrait être la sorte de langage inventé, « en boîte », que les programmeurs transforment en logiciels accessibles aux utilisateurs. Vous vous apprêtez à formater une disquette et votre ordinateur vous « pose » une question amicale : « Êtes-vous sûr de bien vouloir faire ça ? Cela va effacer tout sur le disque ! Répondez Oui ou Non ! » L'utilisateur pensant que l'ordinateur *a l'intention* de lui manifester pareille sollicitude serait très naïf.

Je voudrais donner un peu la parole à un critique. Comme ce

critique imaginaire talonnera nos discussions et investigations dans les chapitres suivants, je lui donnerai un nom : Otto. Donnons-lui la parole :

« C'était un truc facile que d'appeler Shakey "il" plutôt que "cet objet" ; le problème avec Shakey, c'est qu'il n'a pas de véritable intériorité comme nous ; il n'est pas possible de savoir quel effet cela fait que d'être lui. Même si la machinerie qui a enregistré des informations à partir de l'"œil" de sa caméra et qui a transformé cette information en identifications de boîtes avait été très proche de la machinerie qui se trouve dans *notre* système visuel (ce qui n'est pas le cas), et même si la machinerie qui a contrôlé sa production de mots en français avait été très voisine de celle qui se trouve dans le système de production de discours qui contrôle chez nous l'émission de suites de mots français (ce qui n'est pas le cas), quelque chose aurait *encore* manqué : l'intermédiaire au fond de chacun de nous dont les jugements sont exprimés quand nous disons comment cela passe en nous. Le problème avec Shakey, c'est que son entrée et sa sortie sont reliés l'un à l'autre de façon inadéquate : il manque un observateur (celui qui éprouve l'expérience et l'apprécie) se situant quelque part entre l'entrée visuelle et la sortie verbale, en sorte qu'il y ait quelqu'un là-dedans pour *signifier* les mots énoncés par Shakey quand ils sont "dits".

Quand je parle [continue Otto], je signifie ce que je dis. Ma vie consciente est privée, mais je peux choisir de vous en divulguer certains aspects. Je peux décider de vous dire diverses choses au sujet de mon expérience présente ou passée. Quand je fais cela, je formule des phrases que j'adapte avec soin au sujet dont je veux rendre compte. Je peux aller et venir entre l'expérience et le compte rendu proposé, confrontant les mots à l'expérience pour être sûr que j'ai trouvé *les mots justes*¹. Ce vin n'a-t-il pas un petit arrière-goût de *pamplemousse*, ne me rappelle-t-il pas celui des *framboises* ? Faut-il dire que le ton plus élevé avait l'air *plus fort*, ou bien ne serait-ce pas plutôt qu'il m'a semblé être *plus clair* ou *plus distinct* ? Je me concentre sur mon expérience consciente particulière et je parviens à un jugement quant aux mots qui feraient le mieux justice à son caractère. Quand j'ai la satisfaction d'avoir forgé un compte rendu précis, je l'exprime. À partir de mon compte rendu introspectif, vous parvenez à connaître un trait quelconque de mon expérience consciente. »

En tant qu'hétérophénoménologues, il nous faut diviser ce texte en deux parties. Nous devons placer d'un côté les thèses qui portent sur l'expérience qu'éprouve Otto quand il parle. Celles-ci sont inviolables : c'est la façon dont l'expérience est

éprouvée par Otto, et nous devons la prendre comme une donnée qui requiert une explication. À l'autre bout de la chaîne, nous devons placer les thèses théoriques (les conclusions d'argumentations tacites ?) qu'avance Otto au sujet de ce qui se passe en lui et de ce qui se produit dans Shakey, par exemple. Ces thèses n'ont pas de statut particulier, mais nous les traiterons avec le respect que nous devons à toutes les propositions réfléchies.

Je suis tout à fait d'accord pour insister sur le fait que le Petit Homme, l'Observateur Interne du Théâtre Cartésien, doit être éliminé, et non pas découvert, mais nous ne pouvons pas *simplement* le congédier. S'il n'y a pas de Signifieur Central, d'où vient la signification ? Nous devons le remplacer par une analyse plausible de la manière dont une énonciation pourvue de sens – un *vrai* compte rendu, sans guillemets – pourrait être composé sans l'*imprimatur* d'un Signifieur Central solitaire. C'est l'objectif principal de ce chapitre.

Bureaucratie contre Pandémonium

La linguistique contemporaine n'a pas ménagé son intérêt pour la compréhension du langage par un auditeur. Mais elle a, dans une large mesure, ignoré la production du langage par un locuteur. Or c'est en gros la moitié du langage, et la moitié la plus importante de ce point de vue. Bien qu'il existe de nombreuses théories et modèles détaillés de la *perception* du langage et de la *compréhension* des énonciations émises par des locuteurs (les voies qui mènent de la phonologie, à travers la syntaxe, à la sémantique et à la pragmatique) personne – pas même Noam Chomsky, et pas même ses rivaux ou ses disciples – n'a trouvé le moyen de dire quoi que ce soit de très conséquent (même faux) sur les systèmes de *production* du langage. C'est comme si toutes les théories de l'art étaient des théories de l'*appréciation* esthétique ne disant jamais un mot des artistes qui créent – comme si l'art consistait en *objets trouvés*¹ appréciés par les marchands d'art et les collectionneurs.

Il n'est pas difficile de voir pourquoi il en est ainsi. Les énonciations *sont* des objets que nous trouvons facilement pour commencer un certain processus. En fait, la nature des données

brutes ou des entrées d'informations pour les systèmes de perception et de compréhension est tout à fait claire : ce sont des formes ondulatoires de certaines sortes dans l'air, ou des suites de marques sur diverses surfaces planes. Et, bien qu'un certain brouillard obscurcisse les controverses qui portent sur la nature de l'état final du processus de compréhension, ce désaccord profond intervient à la fin du processus étudié, et non au début. Il peut être rationnel d'engager une course quand la ligne de départ est nettement tracée, même si l'on n'est pas tout à fait sûr de l'endroit où elle se terminera. La « sortie » ou le produit de la compréhension du langage sont-ils un *décodage* ou une *traduction* de l'entrée en une nouvelle représentation – une phrase du Mentalais, peut-être, ou une image-dans-la-tête – ou bien un ensemble de *structures profondes*, ou même une entité que nous n'avons pas encore imaginée ? Les linguistes peuvent décider de remettre à plus tard la réponse à ce rébus tout en travaillant sur les parties les plus périphériques du processus.

Avec la production du langage, en revanche, comme personne n'a encore donné de description claire et définitive des origines du processus conduisant finalement à un énoncé complet, il est difficile de simplement ébaucher une théorie. C'est difficile, mais ce n'est pas impossible. Des travaux de qualité ont été réalisés récemment sur les problèmes de la production du langage. Le psycholinguiste néerlandais Pim Levelt les a passés en revue et exposés dans son livre intitulé *Speaking* (1989). Remontant à partir de la sortie, ou encore travaillant vers le centre du processus dans les deux directions, ces recherches donnent un aperçu suggestif de la machinerie qui construit nos énonciations et qui les exprime. (Les exemples qui suivent proviennent de la discussion de Levelt.)

Le discours n'est pas produit par un processus « en bloc » qui supervise et exécute un seul mot à la fois. L'existence d'au moins une capacité limitée de prévision dans le système apparaît dans la manière dont l'accent se trouve distribué dans une énonciation. Voici un cas simple : l'accent dans le mot « dix-neuf » dépend du contexte.

« José : Combien de pesetas ça coûte ?

Miguel : Je pense que c'est dix-NEUF.

José : DIX-neuf pesetas, c'est pas beaucoup. »

Et quand Miguel prononce la seconde parole, il doit ajuster sa prononciation de « dix-neuf » au mot (PESEtas) qui suit. S'il s'était appretté à dire :

« Dix-NEUF, c'est pas beaucoup »

il aurait donné au mot une autre accentuation. Autre exemple : comme l'accent est différent dans les deux occurrences de « Calvados » dans

« Je suis allé en auto de Caen, dans le Calvados, jusqu'à la limite du département du CALVADOS. »

Les contrepèteries et autres erreurs linguistiques montrent très bien comment on observe (et comment on viole) les distinctions lexicales et grammaticales quand on construit un énoncé qu'on va prononcer. Les gens ont plus de facilité à dire « la barre nasale » quand ils veulent dire « l'anal bazar » qu'il n'en ont à dire « la vase bizarre » quand ils veulent dire « la bise avare ». On a tendance à prononcer des mots réels familiers plutôt que les mots simplement prononçables (possibles mais pas réels) même quand la langue fourche. Certaines erreurs indiquent la manière dont doivent opérer les mécanismes de sélection des mots : « Cet individu est un peu sarcastifleur (sarcastique/persifleur²) » et « Je l'ai mis au four à très faible vitesse ». Pensez aussi à la transposition qui doit se produire quand on produit une erreur comme « chanter les maçons » pour « mater l'échanson ».

Grâce à des expériences ingénieuses destinées à provoquer ce genre d'erreurs et à des analyses complexes de ce qui se passe et de ce qui ne se passe pas quand les gens parlent, on a progressé dans l'étude des mécanismes hautement complexes qui articulent un message à partir du moment où *il a été décidé* qu'il fallait produire un certain message. Mais qu'est-ce qui met cette machinerie en marche ? Un lapsus est une erreur : c'est autre chose que ce que le locuteur *avait l'intention de dire*. Quel est le maître d'œuvre qui prescrit ce qu'il y a à faire et relativement à quoi il y a des lapsus comme ceux qu'on vient de voir ?

Qui en est responsable, sinon le Grand Signifieur ? Levelt nous fournit un diagramme, un « plan pour le locuteur » :

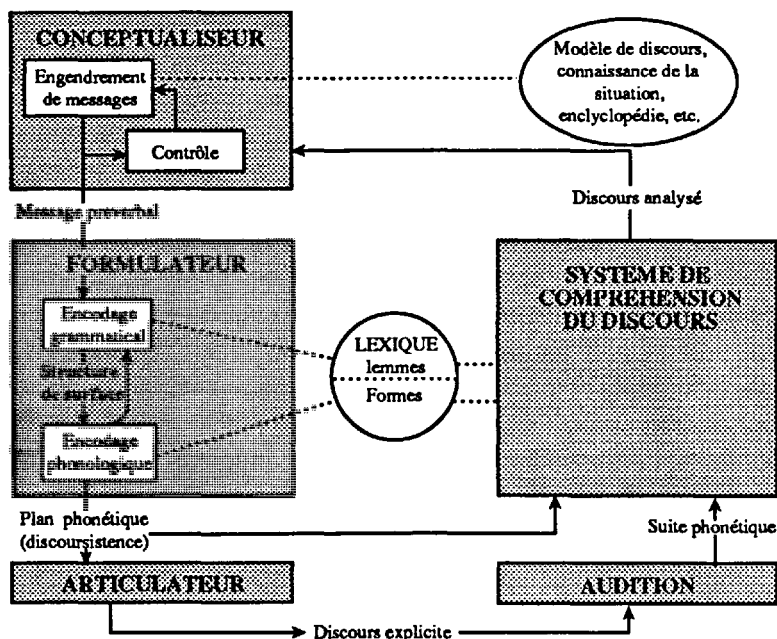


Figure 7.1

De *speaking : From Intention to Articulation*, par Willem J. Levelt, © 1989, Massachusetts Institute of Technology.

Dans le coin supérieur gauche, un fonctionnaire qui a tout l'air d'être le Grand Signifieur est déguisé en Conceptualiseur, armé d'un vaste savoir, de plans et d'*intentions de communication*, et capable d'« engendrer des messages ». Selon Levelt, le Conceptualiseur est « une réification qui doit expliquer plus avant » (p. 9), mais il postule néanmoins son existence, parce qu'il ne peut pas faire démarrer le processus sans faire appel à un Chef non analysé de ce type afin de donner des ordres au reste de l'équipe.

Comment cela marche-t-il ? Le problème sous-jacent sera plus clair si nous commençons par une caricature. Le Conceptualiseur décide d'accomplir un acte de langage, comme insulter son interlocuteur en commentant agressivement la taille de ses pieds. Il envoie donc une commande à la bureaucratie qui dépend de lui, le Département des Relations Publiques (le Formulateur de Levelt) : « Dis à ce mariole que ses pieds sont trop grands ! » Les gens des RP prennent ce travail en charge. Ils trouvent les mots qui conviennent, le pronom possessif à la seconde personne *vos*, un bon mot pour les pieds, et l'adverbe et l'adjectif appropriés. Ils combinent habilement ceux-ci, avec le ton insultant qui convient, et ils passent à l'exécution :

« Tes pieds sont trop grands ! »

Attendez une minute. N'est-ce pas trop facile ? Quand le Conceptualiseur a donné son ordre (ce que Levelt appelle *le message préverbal*), s'il l'a donné en français, comme le suggère ma caricature de ce qui se passe, il a fait tout le travail. Le reste de l'équipe n'a plus guère qu'à émettre le message avec quelques ajustements triviaux. Doit-on dire alors que le message préverbal est formulé dans un autre système représentationnel ou un autre langage ? Quel qu'il soit, il doit être capable de fournir les « éléments » de base à l'équipe de production pour l'objet qu'ils sont supposés composer et émettre, et ceux-ci doivent être formulés dans des termes qu'ils peuvent « comprendre » – pas en français, mais dans une version quelconque de langage cérébral ou de mentalais. Cela devra être une sorte de langage de la pensée, soutient Levelt, mais peut-être est-ce un langage de la pensée qui n'est utilisé que pour produire des actes de langage et qui ne convient pas à toutes les activités cognitives. L'équipe reçoit le message préverbal, un ordre détaillé en mentalais pour produire une énonciation en français, et il émet cet ordre. Cela donne un travail supplémentaire aux subordonnés, mais cela ne fait que masquer la régression qui menace. Comment le Conceptualiseur se figure-t-il la nature des mots en mentalais qu'il doit utiliser pour donner cet ordre ? Il vaudrait mieux qu'il n'y ait pas une réplique à plus petite échelle du schéma de Levelt cachée dans la boîte d'*engendrement du message* (et ainsi de suite à l'infini). Et il ne fait pas de doute que personne n'a dit au Conceptualiseur ce qu'il devait accomplir ; il est, après tout, le Grand Signifieur, l'endroit où la signification *trouve sa source*.

Par conséquent, comment se développe la signification d'un énoncé ? Considérons l'emboîtement suivant d'ordres, qui va de la stratégie globale d'ensemble aux opérations de base en passant par les tactiques spécifiques :

1. Passe à l'offensive !
2. Fais quelque chose de méchant mais de pas trop dangereux pour lui !
3. Insulte-le !
4. Dénigre un aspect de son corps !
5. Dis-lui que ses pieds sont trop grands !
6. Dis : « Tes pieds sont trop grands. »
7. Énonce : Tε pye sô tGo gGâ.

Il ne fait pas de doute que ce réglage dans l'acte final doit se produire. Le discours humain est une activité intentionnelle ; il y a des fins et des moyens, et nous faisons en quelque sorte un travail passable en détaillant les diverses options. Nous aurions pu le bousculer plutôt que l'insulter ; nous aurions pu déprécier son intelligence plutôt que ses pieds ; nous aurions pu aussi citer Fats Waller : « Vos extrémités pédiques sont repoussantes ! »

Mais tout ce réglage est-il accompli par une hiérarchie bureaucratique de commandants donnant des ordres à leurs subordonnés ? Dans cette cascade de commandes, il semble y avoir une bonne quantité de prises de décision – de « moments » où les options sont « sélectionnées » par rapport à leurs rivales. Cela invite à formuler un modèle dans lequel il existe une délégation de responsabilité pour les petits détails et dans lequel les agents subordonnés, qui ont leurs propres intentions, évaluent des raisons de faire diverses sélections. (S'ils ne devaient pas comprendre pourquoi ils font ce qu'ils font, ce ne seraient pas réellement des agents, mais seulement des fonctionnaires passifs, tout juste bons à tamponner les documents qui passent sur leur bureau.)

Le plan de Levelt révèle des traces fossiles de l'une de ses sources : l'architecture de von Neumann qui a été inspirée par les réflexions de Turing sur son propre flux de conscience et qui, en retour, a inspiré bien des modèles en sciences cognitives. Au chapitre 6, j'ai tenté de vaincre la résistance à l'idée selon laquelle la conscience est *plutôt comme* une machine de von Neumann, un processeur sériel reposant sur une succession de contenus définis dévidés à travers le goulot de l'accumulateur. À présent, je dois freiner des quatre fers, et insister sur quelques-unes des façons dont l'architecture fonctionnelle de la conscience humaine n'est *pas* identique à celle d'une machine de von Neumann. Si nous comparons le plan de Levelt à la manière dont les machines de von Neumann émettent habituellement des mots, nous voyons que le modèle de Levelt emprunte trop.

Quand une machine de von Neumann énonce ce qui est écrit dans son cœur, elle exhume les contenus de son unique espace de travail central, l'accumulateur, qui à chaque instant a des contenus entièrement spécifiques dans le langage fixé de l'arithmétique élémentaire. Les « messages préverbaux » rudimentaires d'une machine de von Neumann ont l'allure suivante : 10110101 00010101 11101101. L'une des instructions primitives dans n'importe quel langage machine est une instruction de

SORTIE, qui peut prendre les contenus présents de l'accumulateur (par exemple le nombre binaire 01100001) et les écrire sur l'écran ou sur l'imprimante, en sorte qu'un utilisateur extérieur puisse avoir accès aux résultats accomplis sur l'unité centrale. Selon une variante un petit peu plus proche de l'utilisateur, une opération de routine composée d'une série d'instructions primitives peut d'abord traduire le nombre binaire en notation décimale (par exemple le nombre binaire 00000110 = décimale 6) ou en une lettre de l'alphabet via le code ASCII (par exemple, le 01100001 binaire = « a » et 01000001 = « A »), puis sortir le résultat. Ces sous-routines sont au cœur des instructions de sortie plus complexes qu'on trouve dans les langages de programmation supérieurs, comme Fortran, Pascal ou Lisp. Ces langages permettent au programmeur de créer d'autres sous-routines pour construire des messages plus vastes, en allant chercher de longues séries de nombres dans la mémoire et en les faisant passer à travers l'accumulateur, en les traduisant et en écrivant les résultats sur l'écran ou sur l'imprimante. Par exemple, une sous-routine peut faire plusieurs voyages vers l'accumulateur pour insérer des valeurs dans les blancs dans

Vous avez un découvert sur votre compte, Mr-----, de F
-----, Bonne journée, Mr ----- !

Cette formule « fabriquée » est elle-même stockée sous forme d'une série de nombres binaires dans la mémoire jusqu'à ce qu'une sous-routine quelconque détermine qu'il est temps d'ouvrir la boîte. Ainsi, une hiérarchie stricte de routines fixes peut transformer des suites de contenus spécifiques figurant dans l'accumulateur en expressions qu'un être humain peut lire sur l'écran ou sur l'imprimante : « Voulez-vous sauver ce document ? » ou « 6 fichiers copiés » ou « Salut, Jojo, est-ce que tu veux faire un morpion ? »

Le modèle de Levelt partage deux caractéristiques avec celui-ci : (1) le processus a comme entrée un contenu déjà déterminé et (2) la bureaucratie – le « flux de contrôle » en jargon informatique – doit avoir été planifié avec soin. Toute « prise de décision » s'écoule hiérarchiquement par délégation de responsabilité aux sous-agents dont les profils de poste dictent quelle quantité d'analyse moyen/fin ils sont autorisés à accomplir. Il est intéressant de constater que le premier de ces traits – le contenu déterminé –

Ce verbiage incite les démons à faire d'autres découvertes géniales, auxquelles s'ajoutent une mise en forme qui conduit à d'autres bouts de verbiage plus acceptable, jusqu'à ce que finalement une phrase entière émerge :

Je vais te faire descendre le dentier dans la gorge !

Par chance, cependant, cela reste à l'écart, non parlé, et en même temps (en parallèle) d'autres candidats ont été brassés et se trouvent à présent dans les environs, y compris quelques mauvais candidats évidents comme :

Gros tas !

et

Vous avez lu un livre ces temps-ci ?
Un autre candidat gagne par forfait, c'est lui qui est énoncé :

Tes pieds sont trop grands !

La muse a fait défaut à notre locuteur, en cette occasion ; aucune repartie fine n'a gagné en finale, mais au moins un énoncé correspondant à peu près au « dispositif mental » présent du locuteur s'est trouvé émis. Quand le locuteur s'en ira après la rencontre, il reviendra sans doute sur ce tournoi chaotique, en grommelant et en méditant sur ce qu'il aurait dû dire. Sa muse pourrait alors descendre en lui pour lui souffler quelque chose de meilleur ; le locuteur en goûtera la saveur, retournant ce nouvel énoncé sans cesse dans son esprit, imaginant le regard interloqué qu'il aurait provoqué chez son interlocuteur. Au moment où le locuteur revient à la maison, il peut nettement « se souvenir » avoir embroché son interlocuteur avec une pique bien acérée.

Nous pouvons supposer que tout cela se produit à travers des productions rapides de traitement « gratuit » d'information, avec des hordes de démons anonymes dont les constructions encourageantes ne voient jamais le jour – soit sous la forme d'options qui sont *consciemment* considérées et rejetées, soit comme des actes de langage qui sont finalement exécutés pour être entendus par des étrangers. Si on leur donne assez de temps, plus d'un de ces épisodes peut être essayé silencieusement dans une répétition consciente, mais ce genre d'audition formelle est un évé-

nement relativement rare ; il est réservé pour les occasions où beaucoup de choses sont en jeu et où produire des lapsus peut être très coûteux. Dans le cas normal, le locuteur n'a aucune projection préalable, lui et son auditeur apprennent en même temps ce qu'est l'énonciation du locuteur.

Mais comment ce tournoi verbal est-il jugé ? Quand un mot, une expression ou une phrase entière bat ses rivaux, comment son opportunité ou son adéquation au dispositif mental du moment sont-elles discernées et évaluées ? Qu'est-ce qu'un dispositif mental (sinon une intention de communication explicite), et comment son influence est-elle prise en compte dans le tournoi ? Car après tout, même s'il n'existe pas de Signifieur Central, il doit bien exister une manière pour le contenu de passer du tréfonds du système – des processus perceptifs, par exemple – aux verbalisations.

Passons les problèmes en revue. Le problème posé par l'hypothèse bureaucratique extrême est que le Conceptualiseur semble terriblement puissant : c'est un homoncule qui a trop de savoir et de responsabilité. Cet excès de pouvoir se manifeste dans le problème délicat de savoir comment formuler sa sortie d'information, le *message préverbal*. S'il spécifie déjà un acte de langage – si c'est déjà un acte de langage en mentaux, une commande spécifique envoyée vers le Formulateur – la majeure partie du travail difficile de composition a déjà été faite avant que notre modèle n'intervienne. Le problème posé par l'autre branche de l'alternative, l'hypothèse du Pandémonium, est que nous avons besoin de trouver une façon pour les sources de contenu d'*influencer* ou de *conditionner* les énergies créatrices des démons producteurs de mots sans les leur *dicter*.

Et le processus décrit dans le précédent, les suites de questions qui engendraient des hallucinations dans le modèle du jeu de la psychanalyse ? Rappelons-nous que nous avons éliminé le scénariste freudien des rêves et le producteur d'hallucination en les remplaçant par un processus à partir duquel le contenu *émergeait* sous les interrogations ininterrompues d'un questionneur. Le problème résiduel était de savoir comment se débarrasser du questionneur habile, et nous avons remis ce problème à plus tard. À présent, nous nous trouvons face à un problème complémentaire : comment obtenir des réponses d'un troupeau de concurrents qui nous pressent de questions comme : « Pourquoi ne disons-nous pas " Ta mère a des bottes de militaire ! " » ou

(dans un autre contexte) « Pourquoi ne disons-nous pas : “ Il me semble que je vois un point rouge en mouvement qui devient rouge au fur et à mesure qu’il avance ”. » *Ce sont deux problèmes complémentaires : ne pourrait-on les résoudre en les traitant ensemble ?* Pourquoi ne pas dire que les démons producteurs de mots sont ceux qui répondent et qui jugent ? Des intentions de communication complètes et exécutées – les significations – pourraient émerger d’un processus quasi évolutif de planification d’actes de langage qui implique la coopération, en partie en parallèle, de divers sous-systèmes dont aucun n’est capable à lui seul de produire – ou d’ordonner – un acte de langage.

Un tel processus est-il réellement possible ? Toute une gamme de modèles rendent compte des processus de « satisfaction de conditions », et ils ont en effet des capacités étonnantes d’explication. En plus des diverses architectures « connexionnistes » d’éléments de type neuronal (voir par exemple Mc Clelland et Rumelhart, 1986), il existe d’autres modèles abstraits. L’architecture Jumbo de Douglas Hofstadter (1983), qui chasse des solutions de rébus ou d’anagrammes, a les bonnes caractéristiques, comme les idées de Minsky (1985) au sujet des agents qui composent la « société de l’esprit » – que nous discuterons plus avant au chapitre 8. Mais nous devons suspendre notre jugement jusqu’à ce que soient créés et mis en œuvre des modèles plus détaillés, plus explicites et spécifiquement conçus pour expliquer la production du langage. Il pourra y avoir des surprises et des déceptions.

Nous savons, cependant, que quelque part dans un modèle correct, il nous faudra recourir à un processus évolutionniste d’engendrement de messages. Sans quoi nous aurons sur les bras un miracle (« Et alors un miracle se produit ») ou une régression à l’infini de Signifieurs qui font le travail ³. Nous savons aussi – à partir des recherches analysées par Levelt – qu’il existe des processus très rigides et automatiques qui finissent par l’emporter et qui déterminent les transformations : ils vont de la structure grammaticale à la structure phonologique qui composent la recette musculaire finale conduisant à la parole. Les deux caricatures définissent les extrêmes sur un continu, qui va de l’hyperbureaucratie à l’hyperchaotique. En fait, le modèle proposé par Levelt – par opposition à la caricature que j’ai utilisée pour rendre l’opposition tangible – incorpore (ou peut aisément incorporer) certains des traits non bureaucratiques de la caricature opposée : par exemple, rien de profond ou de structural n’empêche le for-

mulateur de Levelt de s'engager dans une production de langage plus ou moins spontanée ; étant donné la boucle de contrôle qui va du Système de Compréhension du discours au Conceptualiseur (voir figure 7.1), cette activité spontanée *pourrait* jouer un rôle de production semblable à celui qu'on a prêté aux démons producteurs de mots. Entre les deux caricatures, il existe un spectre intermédiaire de façon dont on pourrait développer les autres modèles possibles. La question principale est celle de savoir quel est le degré d'interaction entre les spécialistes qui déterminent le contenu et le style de ce que l'on doit dire d'une part, et ceux qui « connaissent les mots et la grammaire » d'autre part.

À l'un des extrêmes de la chaîne, la réponse est : il n'existe pas d'interaction. Nous pourrions conserver tel quel le modèle de Levelt et lui ajouter simplement un modèle en forme de Pandémonium de ce qui se produit à l'intérieur du Conceptualiseur pour fixer le « message préverbal ». Dans le modèle de Levelt, la séparation est presque complète entre les processus d'engendrement du message (le dispositif qui fait les spécifications) et la production linguistique (le dispositif qui assemble les spécifications). Quand le premier bout de message préverbal arrive au Formulateur, il déclenche la production du début d'une énonciation ; au fur et à mesure que les mots sont choisis par le Formulateur, cela conditionne la manière dont l'énonciation peut continuer, mais il y a une *collaboration* minimale sur la révision des spécifications. Les charpentiers du langage subordonnés qui se trouvent dans le Formulateur sont, comme dit Jerry Fodor, « encapsulés ». À leur façon automatique, ils peuvent faire du mieux qu'ils peuvent avec les ordres qu'ils reçoivent, sans « si », « et » ni « mais ».

À l'autre extrême, on trouve les modèles dans lesquels les mots et les expressions qui viennent du Lexique, avec leurs sons propres, leurs significations et leurs associations, se bousculent au milieu des constructions grammaticales dans un Pandémonium ; tous « essayent » de faire partie du message et certains d'entre eux contribuent par là fortement à la naissance des intentions communicatives qu'un nombre encore plus réduit d'entre eux finissent par exécuter. À cet extrême, les intentions de communication qui interviennent sont autant un effet du processus qu'elles en sont la cause – elles émergent comme un produit, et une fois qu'elles émergent, elles constituent des critères de mesure d'autres implémentations des intentions. Il n'existe pas une seule source de significations, mais beaucoup de sources qui se déplacent,

produites de façon opportuniste par la recherche des mots appropriés. Au lieu qu'un contenu déterminé soit situé dans un lieu fonctionnel particulier, attendant d'être mis en français par des sous-routines, un dispositif mental encore incomplètement déterminé distribué à travers le cerveau conditionne un processus de composition qui démarre d'emblée. Il existe toujours une trame globale de passage en série, concentrée sur un seul sujet à la fois, mais les frontières ne sont pas précises.

Dans le modèle du Pandémonium, le contrôle est usurpé plutôt qu'il n'est délégué, dans un processus qui est largement non planifié et opportuniste : de multiples sources des « décisions » de planification produisent l'énonciation finale, et il n'est pas possible d'opposer les ordres de marche des contenus qui s'écoulent de l'intérieur et les suggestions volontaires d'implémentation établies par les démons producteurs de mots. Ce type de modèle suggère que, pour *préserver* le rôle créateur du dispositif d'*expression* des pensées (quelque chose qui comptait beaucoup pour Otto), il nous faut *abandonner* l'idée que le dispositif qui *pense* les pensées part d'une pensée déterminée qu'il entend exprimer. Cette idée d'un contenu déterminé comptait aussi beaucoup pour Otto, mais il faut la laisser tomber (la quatrième section de ce chapitre explorera les diverses possibilités plus en détail).

Entre ces deux extrêmes, où est la vérité ? C'est une question empirique dont nous ne connaissons pas encore la réponse⁴. Certains phénomènes, pourtant, suggèrent fortement (à mes yeux du moins) que la production du langage implique un Pandémonium – des processus opportunistes, parallèles, évolutifs – presque à toutes les étapes. On les passera en revue rapidement dans la section qui suit.

Quand les mots veulent se faire dire

« Quoi que nous voulions dire, nous ne disons probablement pas exactement *cela*. »

Marvin Minsky, 1985, p. 236.

Les chercheurs en intelligence artificielle Lawrence Birnbaum et Gregg Collins (1984) ont fait remarquer que les lapsus freu-

diens ont une particularité. On sait que Freud a attiré notre attention sur les lapsus qui ne sont pas aléatoires ou dénués de sens, mais selon lui, profondément significatifs : il s'agit d'insertions inconsciemment *intentionnelles* dans le corps du discours, qui satisfont indirectement ou partiellement des objectifs de communication réprimés du locuteur. Les sceptiques ont souvent rejeté avec violence cette thèse freudienne usuelle. Mais son application à des cas particuliers pose un problème qui n'a rien à voir avec ce qu'on peut penser des thèmes freudiens plus obscurs portant sur la sexualité, le complexe d'Œdipe ou les désirs de mort. Freud discute un exemple dans lequel un individu dit :

« Messieurs, je rote à la santé de notre Directeur. »

(En allemand – la langue employée dans cet exemple – le mot « roter » (*aufzustossen*) a glissé à partir du mot « boire » (*anzus-tossen*).)

« Dans son explication, Freud soutient que ce lapsus est une manifestation d'un objectif inconscient de la part du locuteur de ridiculiser ou d'insulter son supérieur, réprimé par le devoir social et politique de lui faire honneur. Cependant [...] on ne peut raisonnablement s'attendre à ce que l'intention du locuteur de ridiculiser son supérieur ait initialement donné naissance à un plan d'utilisation du mot " roter " : *a priori*, il existe des centaines de mots et d'expressions qui pourraient être utilisées de façon plus plausible pour insulter ou pour ridiculiser quelqu'un... On ne voit pas en quoi un processus planificateur aurait pu raisonnablement anticiper que le but de ridiculiser ou d'insulter son supérieur pourrait être satisfait en énonçant le mot " roter ", pour exactement la même raison pour laquelle il est improbable que le planificateur ait utilisé le mot comme une insulte en premier lieu. »

Le seul processus qui pourrait expliquer la fréquence des heureux hasards que sont les lapsus freudiens, selon eux, est un processus de « planification opportune ».

« Ce que des exemples comme le précédent semblent indiquer, par conséquent, est que les buts *eux-mêmes* sont des agents cognitifs actifs, capables de commander les ressources cognitives nécessaires pour reconnaître des opportunités susceptibles de les satisfaire, et les ressources comportementales requises pour utiliser les opportunités » (Birnbaum et Collins, 1984, p. 125).

Les lapsus freudiens se signalent par le fait qu'ils semblent être des erreurs et en même temps ne pas en être, mais le fait (si c'en est un) qu'ils satisfont des objectifs *inconscients* ne les rend pas plus difficiles à expliquer que d'autres choix de mots qui jouent plusieurs rôles (ou satisfont plusieurs objectifs) en même temps. Il est à peu près aussi difficile d'imaginer comment des jeux de mots et d'autres formes d'humour verbal pourraient être le résultat de planifications et de productions non opportunistes, encapsulées. Si quelqu'un dispose d'un plan pour concevoir des mots d'esprit – un plan détaillé qui marche effectivement – plus d'un comédien serait prêt à payer pour l'avoir ⁵.

Si Birnbaum et Collins ont raison, l'usage créatif du langage n'est possible que par l'intermédiaire d'un processus parallèle dans lequel de multiples *objectifs* sont simultanément en alerte pour utiliser différents matériaux. Mais que dire si les matériaux eux-mêmes étaient en même temps en alerte à la recherche d'opportunités à incorporer ? Nous sélectionnons notre vocabulaire à partir de notre culture ; les mots et les expressions sont les traits phénotypiques les plus saillants – les corps visibles – des mêmes qui nous envahissent, et il pourrait difficilement exister un dispositif plus propre à faire se répliquer les mêmes que le système de production de langage dans lequel les bureaucrates chargés de superviser le tout auraient en partie abdiqué, cédant une bonne partie de leur contrôle aux mots eux-mêmes, qui en effet se battent entre eux pour avoir une chance de se produire en public.

Ce n'est pas une nouveauté : nous disons une partie de ce que nous disons parce que cela sonne à nos oreilles d'une certaine façon, et non parce que nous aimons ce que cela signifie. Un nouvel argot se répand dans certaines sous-communautés, trouvant lentement son chemin dans le discours de presque tout le monde, même ceux qui essaient d'y résister. Peu de gens parmi ceux qui utilisent un nouveau mot suivent délibérément ou consciemment la maxime du maître d'école : « Utilisez un mot trois fois et il est à vous ! » À des niveaux plus vastes d'agrégation, des phrases entières nous plaisent pour la façon dont elles sonnent à nos oreilles ou dont elles apparaissent sur notre langue, tout à fait indépendamment de la question de savoir si elles rencontrent des spécifications propositionnelles que nous aurions déjà décidées. L'une des phrases d'Abraham Lincoln qui se prêtent le plus à citation est :

« Vous pouvez tromper tout le monde quelques fois, et quelques personnes tout le temps, mais vous ne pouvez pas tromper tout le monde tout le temps ⁶. »

Que voulait dire Lincoln ? Les professeurs de logique aiment à faire remarquer qu'il y a « une ambiguïté de portée » dans la phrase. Lincoln entendait-il affirmer qu'il existe quelques naïfs qu'on peut toujours tromper, ou qu'en toute occasion, une personne ou une autre sera trompée – mais pas toujours les mêmes personnes ? Logiquement parlant, ce sont des propositions entièrement différentes.

Comparons :

« Quelqu'un gagne toujours à la loterie »

« Y a un truc ! »

« Ce n'est pas ce que je voulais dire. »

Quelle était l'interprétation visée par Lincoln ? Peut-être ni l'une ni l'autre ! Quelles sont les chances pour que Lincoln n'ait jamais noté l'ambiguïté de portée et n'ait jamais entendu avoir l'une des intentions communicatives plutôt que « l'autre » ? Peut-être cela a-t-il sonné à ses oreilles comme un si bon mot qu'il n'a jamais vu l'ambiguïté, *et n'a jamais eu d'intention de communication préalable* – sauf celle de dire de quelque chose de lapidaire et de bien cadencé sur le thème général de la mise en boîte. Les gens parlent en fait comme cela, même les grands Signifieurs comme Lincoln.

La romancière Patricia Hampl, dans un essai subtil intitulé *Les Habitudes relâchées de l'imagination libre*, parle de sa propre technique de composition des nouvelles

« Toute histoire a une histoire. Cette histoire secrète, qui a peu de chances d'être racontée, est l'histoire de sa création. Peut-être que " l'histoire de l'histoire " ne *peut* jamais être racontée, parce qu'un travail une fois terminé consume sa propre histoire, la rend obsolète, en fait une coque vide » (Hampl, 1989, p. 37).

Le travail abouti, note-t-elle, peut immédiatement être interprété par les critiques comme un artefact habilement conçu pour réaliser une quantité d'intentions sophistiquées de l'auteur. Mais quand elle envisage ces hypothèses au sujet de son propre travail, elle est dans l'embarras :

« " Hampl " avait quelques précieuses intentions, sauf, comme le charlatan que je me suis soudain sentie incarner, celle de chiper

tout ce qui traînait sur la table qui puisse servir mes objectifs immédiats. Pire, les "objectifs" étaient vagues, incohérents, réversibles, sous pression. Et qui, ou quoi, pouvait bien appliquer la pression ? Je ne pouvais pas le dire » (p. 37).

Comment fait-elle alors ? Elle suggère une maxime : « Content-toi de continuer à parler – il suffit de marmonner. » Finalement, le marmonnement prend des formes que l'auteur approuve. Se pourrait-il que le processus qu'Hampl détecte à grande échelle dans son travail créatif d'écriture ne soit qu'une expansion des processus sous-jacents et rapides qui produisent le discours créatif dans la vie quotidienne ?

Cette similarité tentante n'implique pas seulement un processus, mais aussi une attitude ou une réaction ultérieure. Le zèle mis par Hampl à se confesser tranche par rapport à la réaction normale – et pas vraiment malhonnête – des auteurs face aux interprétations amicales de leurs lecteurs : ils s'en remettent aimablement à toute imputation d'intention, et ils sont même prêts à les approfondir. Cela donne des remarques du style : « Je suppose que c'était bien ce que je voulais dire, dès le départ ! » Et pourquoi pas ? Est-il incohérent de penser qu'une décision qu'on vient de prendre (aux échecs, dans la vie, dans l'écriture) est en fait plus habile qu'on ne l'avait réalisé au départ ? (Pour d'autres réflexions sur ce sujet, voir Eco, 1990.)

Comme dit E.M. Forster : « Comment puis-je savoir ce que je pense avant de voir ce que je dis ? » Nous découvrons souvent ce que nous pensons en réfléchissant à ce que nous nous découvrons en train de dire – et non de corriger. Nous sommes donc, au moins dans ces occasions, embarqués sur le même bateau que nos critiques et nos interprètes qui, rencontrant un morceau de texte, en font la meilleure lecture qu'ils puissent trouver. Le fait que nous l'ayons dit lui donne un certain pouvoir de conviction, ou au moins une présomption d'authenticité. *Probablement*, si je l'ai dit (et je me suis entendu moi-même le dire, et je ne me suis pas entendu moi-même proposer immédiatement des amendements), je veux le dire ; cela veut dire probablement ce que ce cela veut dire – pour moi.

On trouve un bon exemple de cela dans la vie de Bertrand Russell :

« Les deux hôtes partirent fort tard et Russell resta seul avec Lady Ottoline. Ils s'assirent devant le feu jusqu'à quatre heures du

matin. Russell, se rappelant la scène quelques jours plus tard, écrivit : " Je ne savais pas que je vous aimais jusqu'à ce que je me sois entendu vous le dire – pendant un instant je pensai : ' Dieu tout-puissant, qu'ai-je dit là ? ' et je sus alors que c'était la vérité. " » (Clark, 1975, p. 176.)

En d'autres occasions, cependant, nous n'avons pas le sentiment de faire ce genre de *découverte* auto-interprétative. Nous pourrions supposer que dans ces cas, les cas normaux, nous avons un pressentiment intime et privilégié de ce que nous voulons dire, simplement parce que nous sommes nous-mêmes les Signifieurs, la *fons et origo* du sens des mots que nous prononçons. Mais c'est une supposition qu'il faut justifier, et pas seulement accepter parce qu'elle est traditionnelle. Car il pourrait bien se faire que la raison pour laquelle nous n'avons aucun sentiment de découverte dans ces cas soit que ce que nous voulons dire nous apparaisse évident. Je n'ai pas besoin de sentiment d'« accès privilégié » quand je dis « s'il vous plaît, passez-moi le sel » à table : je demande qu'on me passe le sel.

J'ai cru un moment que la seule solution consistait à postuler un Signifieur Central, mais je pensais que je lui avais trouvé un port où il serait en sécurité. Dans *Content and Consciousness*, j'ai soutenu qu'une ligne de division fonctionnellement nette (que j'appelais la ligne de partage de la conscience (*awareness*)) devait séparer la fixation préconsciente des intentions de communication de leur exécution ultérieure. La localisation de cette ligne dans le cerveau était peut-être horriblement difficile à tracer, du point de vue anatomique, mais il fallait qu'elle existe, logiquement, comme ligne permettant de partager les dysfonctionnements en deux espèces. Des erreurs pouvaient intervenir n'importe où dans l'ensemble du système, mais toute erreur devait – par une nécessité géométrique – tomber d'un côté de la ligne ou de l'autre. Si elles tombaient du côté de l'exécution, c'étaient des erreurs (corrigibles) d'*expression*, comme les lapsus *linguae*, les pataquès, les erreurs de prononciation. Si elles tombaient du côté interne ou plus élevé de la ligne, *elles changeaient ce qui devait être exprimé* (le message préverbal, dans le modèle de Levelt). La signification se fixait à partir de cette ligne de partage : je pensais que c'était de là qu'elle venait, parce que *quelque chose* doit jouer le rôle de critère par rapport auquel le « feedback » puisse enregistrer l'incapacité d'exécuter l'acte de langage.

Mon erreur consistait à être victime du même genre d'am-

biguité de portée que celle qui emprisonne l'interprétation de la formule de Lincoln. Il doit bien exister quelque chose à chaque occasion qui soit, pour la circonstance, le critère par rapport auquel toute « erreur » corrigée soit corrigée, mais il n'est pas nécessaire que ce soit la même chose à chaque fois – même *pendant* la durée d'un seul acte de langage. Il n'est pas nécessaire qu'il existe une ligne (même mal tracée) *fixe* pour marquer cette distinction. En fait, comme nous l'avons vu au chapitre 4, la distinction entre des révisions préalables à l'expérience qui *changent ce dont on a fait l'expérience* et des révisions postérieures à l'expérience qui ont pour effet *de rendre compte faussement ou de mal enregistrer ce dont on a fait l'expérience* est indéterminée. Les sujets sont quelquefois amenés à réviser ou à amender leurs assertions, et quelquefois ils ne sont pas amenés à le faire. Il arrive que, quand ils font effectivement des révisions, le récit édité ne soit pas plus près de « la vérité » ou de « ce qu'ils voulaient *réellement* dire » que la version antérieure qu'il remplaçait. Comme nous l'avons noté auparavant, la distinction entre l'endroit où cesse la prépublication et l'endroit où interviennent la post-publication et l'insertion des errata ne peut être tracée qu'arbitrairement. Quand nous demandons à un sujet si son aveu public représente la vérité intime et ultime de ce qu'il vient d'éprouver, il n'est pas en meilleure position pour en juger que nous le sommes nous-mêmes, en tant qu'observateurs extérieurs (voir aussi Dennett, 1990d).

On peut envisager ce même phénomène sous un autre angle. Chaque fois que le processus de création d'une expression verbale se produit, il y a dès le départ une distance à éliminer : « la distance de non-correspondance dans l'espace sémantique », comme nous pourrions l'appeler, entre le contenu qui est en position d'être exprimé et les divers candidats à l'expression verbale qui se trouvent appelés en premier. (Dans mon ancienne théorie, je traitais cela comme un problème de simple « correction de feedback », avec un *point fixe* pour jouer le rôle de critères par rapport auquel mesurer, rejeter, améliorer les divers candidats.) Le processus de va-et-vient qui réduit la distance est une sorte de feedback, mais il se peut tout aussi bien que le contenu-à-exprimer soit ajusté dans le sens d'une expression candidate, ou que cette expression candidate soit remplacée ou éditée de manière à pouvoir s'adapter au contenu-à-exprimer. C'est ainsi que les mots ou tournures les plus accessibles ou les plus dis-

ponibles pourraient *changer le contenu de l'expérience* (si nous comprenons l'expérience comme ce qui se trouve en dernière instance rapporté – l'événement établi dans le monde hétérophénoménologique du sujet) ⁷.

Si notre unité en tant que Signifieurs n'est pas plus garantie que cela, alors elle devrait en principe pouvoir être mise en pièces en certaines occasions. Voici deux cas dans lesquels il semble que cela ce soit produit.

J'ai eu un jour la malchance de me retrouver premier arbitre de base dans une partie de base-ball – tâche nouvelle pour moi. À un moment décisif pendant la partie, c'est à moi qu'a incombé la responsabilité de décider du statut du batteur qui devait courir en premier. Il fallait l'appeler vite, et je me suis retrouvé en train de lever le pouce emphatiquement vers le haut – le signal pour « OUT » – tout en criant « BON ». Dans le tumulte qui s'ensuivit, on me pressa d'expliquer ce que j'avais voulu dire. Honnêtement, j'en étais incapable, du moins selon un point de vue privilégié. Je décidai finalement (pour moi) que, puisque je n'avais pas la pratique des signaux manuels mais que j'étais un locuteur compétent, il fallait suivre mon acte vocal. N'importe qui en aurait jugé de même. (Je serais heureux d'apprendre d'autres anecdotes dans lesquelles les gens ne savent pas lequel des deux actes de langage différents ils ont voulu produire.)

Dans une situation expérimentale, le psychologue Tony Marcel (sous presse) a découvert un cas encore plus dramatique. On demanda à un sujet qui souffrait de vision aveugle (nous le retrouverons au chapitre 11) de *dire* quand il pensait que se produisait un éclair lumineux. Mais on lui donnait des instructions particulières quant au moment où il était supposé faire cela. On lui donnait l'instruction de produire cet acte de langage par l'intermédiaire de trois actes distincts en même temps (pas à la suite, mais pas nécessairement « à l'unisson » non plus) :

1. Dire « Oui »
2. Presser un bouton (le bouton pour OUI)
3. Cligner de l'œil pour OUI.

Fait étonnant, le sujet n'accomplissait pas toujours les trois choses ensemble. Certaines fois, il clignait de l'œil pour OUI mais ne disait pas OUI ni ne poussait le bouton OUI, et ainsi de suite. Cependant, il était impossible de ranger les trois réponses dis-

tinctes dans un ordre quelconque, selon la fidélité par rapport aux intentions, ou par rapport au degré de précision. Autrement dit, quand les trois actions ne s'accordaient pas, le sujet n'avait pas de guide qui lui permette de dire quel acte il fallait accepter et lequel devait compter comme un lapsus verbal, un lapsus manuel ou un lapsus des paupières.

Il reste à voir si l'on peut aboutir à des résultats semblables dans d'autres conditions et avec d'autres sujets. Mais il existe d'autres conditions pathologiques dans lesquelles on *peut* enclencher une verbalisation sans qu'un Signifieur Central donne l'ordre de mise en marche. Si l'on souffre de ces pathologies, « l'esprit est en vacances, mais la bouche fait des heures supplémentaires », comme le dit la chanson de Mose Allison.

L'*aphasie* est une perte ou le résultat d'une lésion de l'aptitude à parler. Il en existe plusieurs sortes, qui sont très fréquentes et que les neuropsychologues et les linguistes ont beaucoup étudiées. Dans le cas le plus courant, l'aphasie de Broca, le patient est bien conscient du problème et lutte, avec un sentiment croissant de frustration, pour trouver les mots qu'il a sur le bout de la langue. L'existence d'intentions communicatives inabouties est alors douloureusement et clairement ressentie par le patient. Mais il existe une autre sorte d'aphasie, relativement rare : l'aphasie du jargon, dans laquelle les patients ne semblent pas angoissés de leur déficit verbal⁸. Bien que leur intelligence soit normale et qu'ils ne soient en rien psychotiques ou déments, ils semblent entièrement se satisfaire de performances verbales comme les suivantes (empruntées à deux cas décrits par Kinsbourne et Warington, 1963) :

Cas 1

Comment ça va aujourd'hui ?

– Comméragage d'accord et Lords et le cricket et les batailles d'Angleterre et d'Écosse. Je ne sais pas. L'hypertension et deux ont gagné au cricket, servant, donnant des coups de batte et attrapant, pauvres vieux, les annulations peuvent être du comméragage, les annulations, bras et argument, finissant de servir.

– Quelle est la signification de « sécurité d'abord » ?

– Regarder et voir et la Route de Richmond en particulier, et regarder la circulation et hésiter à droite et se promener, très bonne cause, peut-être, les zèbres peut-être ceux-là, la voiture et le feu rouge.

Cas 2

Travaillez-vous dans un bureau ?

– Je travaillais en effet dans un bureau.

– *Et pour quelle entreprise ?*

– Oh, comme cadre, et le sujet de plainte et de discuter les tonations pour savoir quel type elles étaient, et comment elles étaient imprimées et écartées du différent... tricu... tricola, pour m'éviter les conventements attribués... Pardon...

– Elle veut donner à quelqu'un la vocation subjective de maintenir la vocation de l'imprégnation parfaite de consanguinité.

– Sa corrucation normale serait un point.

On lui demande d'identifier une lime à ongles :

– C'est un couteau, une queue de couteau, un vieux, vieux, couteau.

Et des ciseaux :

– Des bosquets – c'est des bosquets – c'est pas vraiment des bosquets – deux bosquets contenant un peigne – non, pas un peigne – deux bosquets à condition que le commandant ne soit pas maintenant.

Un cas étrangement similaire, et bien plus commun, est l'*af-fabulation*. Au chapitre 3, j'ai suggéré que les gens normaux pouvaient souvent fabuler sur les détails de leur propre expérience, parce qu'ils ont tendance à inventer sans s'en apercevoir, et qu'ils confondent théoriser et observer. L'affabulation pathologique est une fiction involontaire d'un type entièrement différent. Il arrive souvent, dans les cas de lésions cérébrales, en particulier quand les gens ont des pertes de mémoire terribles – comme dans le syndrome de Korsakoff (une conséquence typique de l'alcoolisme grave) – qu'ils se mettent à déblatérer des choses totalement fausses sur leur vie ou sur leur histoire passée, et même au sujet des événements qui se sont produits pendant les cinq dernières minutes, si leur amnésie est grave.

Le verbiage qui en résulte a une allure pratiquement normale. Il sonne souvent, en fait, comme le bavardage sans conséquences, stéréotypé, qui passe pour de la conversation dans les bars : « Oh oui, ma femme et moi, nous avons vécu dans la même maison pendant trente ans – j'allais souvent à Coney Island, et, vous

savez, je m'asseyais sur la plage — *J'adorais* m'asseoir sur la plage, je regardais juste les jeunes, et c'était avant l'accident... » Sauf que c'est totalement inventé. La femme en question peut être morte des années auparavant, n'avoir jamais vécu à moins de quatre cent kilomètres de Coney Island, et ils peuvent avoir déménagé d'appartement en appartement. Un auditeur non initié peut souvent ne pas se rendre compte qu'il est en face d'un affabulateur, tant les réminiscences et les réponses immédiates aux questions sont chez lui naturelles et « sincères ».

Les affabulateurs ne se rendent pas compte qu'ils inventent de toutes pièces, et les aphasiques du jargon oublient qu'ils déclament des salades. Ces anomalies étourdissantes sont des cas d'*anosognosie*, ou d'incapacité à reconnaître un déficit. Il y a d'autres sortes d'absence de contrôle de soi de ce genre, et au chapitre 10 nous verrons ce qu'elles peuvent nous apprendre sur l'architecture fonctionnelle de la conscience. En attendant, nous pouvons noter que la machinerie du cerveau est tout à fait capable de construire des actes de langage apparents en l'absence d'une quelconque direction cohérente venue d'en haut ?.

La pathologie, sous la forme de la tension temporaire induite par des expérimentateurs habiles ou des dépressions plus permanentes causées par la maladie ou des accidents mécaniques dans le cerveau, nous donne quantité d'indices sur l'organisation de la machinerie. Ces phénomènes suggèrent que notre seconde caricature, le Pandémonium, est plus proche de la vérité que le modèle bureaucratique, plus digne, mais nous devons encore nous en assurer empiriquement. Je ne soutiens pas qu'un modèle dans l'ensemble bureaucratique ne puisse pas rendre justice à ces pathologies, mais seulement qu'elles ne semblent pas être des dysfonctionnements naturels de ce genre de système. Dans l'Appendice B, pour les savants, je mentionnerai quelques directions de recherches qui pourront nous aider à confirmer ou à infirmer mon intuition.

Dans ce chapitre, j'ai esquissé — sans prétendre le prouver — la manière dont un torrent de produits verbaux émergeant de milliers de démons producteurs de mots formant des coalitions temporaires pouvait manifester une unité, l'unité d'une interprétation optimale et progressive, qui fait apparaître ces produits *comme s'ils* étaient les intentions réalisées par un Conceptualiseur. Et c'est bien ce qu'elles sont, mais sans être les intentions d'un Conceptualiseur *interne* qui fasse partie du système de pro-

duction du langage. Ce sont les intentions d'un Conceptualiseur global, la personne, dont le système de production du langage fait partie.

Cette idée peut paraître de prime abord étrange, mais elle n'a rien de surprenant. En biologie, nous avons appris à résister à la tentation d'expliquer *l'organisation dans les organismes* en postulant l'existence d'une unique grande Intelligence qui ferait tout le travail. En psychologie, nous avons appris à résister à la tentation d'expliquer *la vision* en disant que tout se passe comme s'il y avait un visionneur d'écran interne, parce que le visionneur d'écran fait tout le travail – la seule chose qu'il y ait entre un tel homoncule et les yeux est une sorte de câble de télé. Nous devons résister également à la tentation d'expliquer *l'action* comme le produit des impératifs d'un commandant interne des actions qui fait une trop grande partie du travail de planification. Comme d'habitude, la bonne façon de se débarrasser d'une intelligence trop grande pour notre théorie consiste à la remplacer par une fabrique en dernière instance mécanique de semi-intelligences semi-indépendantes qui agissent de concert.

Ces remarques ne s'appliquent pas seulement à la production du discours. Elles s'appliquent systématiquement à l'action intentionnelle (voir Pears, 1984, pour des idées voisines). Et contrairement aux apparences, la phénoménologie nous aide à voir qu'il en est ainsi. Bien que nous soyons à l'occasion conscients d'effectuer des raisonnements pratiques complexes, conduisant à une conclusion quant à ce que, tout bien considéré, nous devrions faire, suivis d'une décision consciente de faire la chose en question, et aboutissant finalement à l'action correspondante, nous n'éprouvons en fait que rarement cela. La plupart de nos actions intentionnelles sont accomplies sans ce type de préambule, et c'est une bonne chose, parce que sinon nous n'aurions pas le temps. On se fourvoie souvent parce qu'on suppose que les cas relativement rares de raisonnement pratique conscient peuvent servir de modèle pour les autres, les cas dans lesquels nos actions intentionnelles émergent de processus auxquels nous n'avons pas accès. Nos actions nous satisfont, en règle générale ; nous reconnaissons qu'elles sont, dans l'ensemble, cohérentes, et qu'elles contribuent de façon adéquate et synchrone à nos projets, tels que nous les comprenons. C'est pourquoi nous supposons en confiance qu'elles résultent de processus qui sont sensibles de façon fiable aux fins et aux moyens. C'est-à-dire qu'elles sont

rationnelles, en *un* certain sens (Dennett, 1987a, 1991a). Mais cela ne veut pas dire qu'elles sont rationnelles en un sens plus étroit : elles ne sont pas le produit d'un raisonnement suivi. Nous n'avons pas à expliquer les processus internes quand nous adoptons le modèle qui postule un agent interne qui raisonne, conclut, décide et qui adapte méthodiquement les moyens aux fins et ordonne les actions planifiées. Nous avons vu en gros comment un processus de type différent pourrait contrôler le discours, aussi bien que nos autres actions intentionnelles.

Lentement, mais sûrement, nous commençons à abandonner nos mauvaises habitudes de pensée, et à les remplacer par de meilleures habitudes. La déposition du Signifieur Central est plus généralement une déposition de l'Intentionneur Central, mais le Chef peut toujours se retrouver sous d'autres déguisements. Au chapitre 9, nous le rencontrerons dans les rôles de l'Observateur et du Rapporteur, et il nous aidera à trouver d'autres façons de penser ce qui se passe. Mais il nous faut d'abord assurer les fondements de nos nouvelles habitudes de pensée en les liant plus étroitement à quelques précisions scientifiques.

NOTES

1. En français dans le texte (NdT).

2. Mot-valise emprunté à Boris Vian (NdT).

3. Dan Sperber et Deirdre Wilson (1986) proposent une nouvelle perspective sur la manière dont nous composons nos communications en insistant sur des modèles expliquant la façon dont les choses *marquent* effectivement, dans le locuteur et dans l'auditeur, contrairement à la pratique courante des philosophes et des linguistes qui ont tendu à se contenter d'allusions au sujet des mécanismes tout en faisant appel à des reconstructions rationnelles des tâches supposées et de ce qu'elles exigent. Cela permet à Sperber et Wilson de considérer des problèmes de possibilités pratiques et d'efficacité : des principes du moindre effort, et des prises en compte du temps et de la probabilité. Ils montrent alors, selon cette perspective, comment certains « problèmes » traditionnels disparaissent, en particulier le problème de savoir comment l'auditeur trouve la « bonne » interprétation de ce que le locuteur a eu l'intention de communiquer. Bien qu'ils ne rattachent pas leur modèle au niveau des processus évolutifs du type de ceux que j'ai considérés, celui-ci invite certainement ce genre de complément.

4. Comme le note Levelt : « Si l'on pouvait, par exemple, montrer que cet engendrement de messages est directement affectée par l'accessibilité des lemmes ou de formes verbales, on pourrait établir qu'il y a un feedback direct du Formulateur au Conceptualiseur. C'est une question empirique, qu'il est possible de vérifier... Pour le moment, on n'a pas de preuve qu'il existe un tel feedback » (p. 16). Les données qu'il passe en revue proviennent d'expériences très contrôlées dans lesquelles on donne au locuteur une tâche spécifique à effectuer : comme *décrire l'image sur l'écran aussi vite que vous pouvez* (p. 276-282). C'est un test négatif excellent – pour ma part j'ai été surpris de l'absence d'effets dans ces

expériences – mais, comme il le reconnaît, ce n'est pas concluant. On ne fait pas de déclaration *ad hoc* quand on dit que le caractère artificiel de ces situations expérimentales a réussi à noyer la dimension créatrice et opportuniste de l'usage du langage. Levelt a raison : peut-être que le seul feedback du Formulateur au Conceptualiseur est *indirect* : c'est la sorte de feedback qu'une personne ne peut produire *qu'en* se parlant explicitement à elle-même et en se forgeant une opinion sur ce qu'elle se trouve en train de dire.

5. Levelt me dit qu'il est lui-même un chasseur de jeux de mots invétéré et il sait très bien comment il procède : « En m'entraînant depuis des lustres, je tourne autour de tout mot que j'entends. Puis (tout à fait consciemment) je vérifie le résultat pour voir s'il a un sens. Dans 99,9 % des cas, rien d'amusant n'en sort. Mais un sur un millier me suffit, et ces jeux de mots-là, je les exprime immédiatement (communication personnelle). C'est un exemple parfait de résolution de problèmes à la von Neumann : sériel, contrôlé, et *conscient* ! La question est de savoir s'il y a d'autres façons, plus pandémoniques, d'engendrer des mots d'esprit inconsciemment.

6. Selon l'*Oxford Dictionary of Quotations* (2^e édition, 1953), cette phrase célèbre est aussi attribuée à Phineas T. Barnum. Comme Barnum est un illustre et généreux ancien élève de mon université, je sens de mon devoir d'attirer l'attention sur la possibilité que Lincoln n'ait pas été à l'origine de ce même hautement répliatif.

7. Cela rappelle la conception freudienne du « préconscient » : « La question : « comment une chose devient-elle consciente ? » serait mieux posée ainsi : « Comment une chose devient-elle préconsciente ? », et la réponse serait : « En étant reliée aux présentations verbales qui lui correspondent » » (*Le moi et le ça*, 1962, p. 10).

8. Levelt me dit qu'il y a des recherches en cours à l'Institut Max Planck car les psycholinguistes à Nimègue mettent en doute ce point, c'est-à-dire la thèse officielle. Des travaux de Heeschen suggèrent que certains aphasiques du jargon ou de Wernicke soient bien angoissés par leur déficit, et semblent adopter une stratégie de répétition, dans l'espoir de parvenir à communiquer.

9. Un autre phénomène linguistique anormal est le symptôme connu de la schizophrénie : « entendre des voix ». Il est à présent bien établi que la voix qu'entend le schizophrène est la sienne : il se parle en silence à lui-même sans le réaliser. Un obstacle aussi simple qu'obliger le patient à ouvrir grand sa bouche suffit pour supprimer les voix (Bick et Kinsbourne, 1987). Voir aussi Hoffman (1986) et le commentaire d'Akins et Dennett, « Who may I say is calling ? » (1986).

Chapitre 8

L'architecture de l'esprit humain

Où en sommes-nous ?

La partie la plus dure de notre travail est derrière nous, mais il nous reste encore beaucoup à faire. Nous avons à présent accompli les exercices les plus ardues d'extension de nos capacités d'imagination, et nous sommes prêts à mettre en pratique notre nouvelle perspective. En chemin, nous avons laissé plusieurs problèmes sans réponse, et nous nous sommes autorisé beaucoup d'approximations. Il y a des promesses à tenir, et il nous faut encore préciser les thèses et les comparaisons que nous avons différées. La théorie que j'ai développée inclut des éléments empruntés à divers penseurs. J'ai quelquefois ignoré délibérément ce que ces penseurs considèrent comme les parties les plus fortes de leurs théories, et j'ai mêlé des idées provenant de divers camps « hostiles » à mes propres vues, mais j'ai supprimé ces détails embrouillés pour plus de clarté et de netteté. Certains théoriciens sérieux de l'esprit peuvent s'en être trouvés frustrés, mais je ne voyais pas comment procéder autrement pour laisser différents types de lecteurs saisir le nouveau point de vue. À présent, cependant, nous sommes en bonne position pour faire le point et pour fournir quelques détails essentiels. Après tout, la raison pour laquelle on doit se fatiguer à construire une nouvelle perspective est que cela nous permet d'envisager les phénomènes différemment. Faisons donc un bref tour d'horizon.

Sous forme de croquis schématique, voici ma théorie telle qu'elle apparaît jusqu'à présent :

Il n'existe pas de « flux de conscience » unique, définitif, parce qu'il n'y a pas de Quartier Général central, pas de Théâtre Cartésien « où tout se rejoint » pour le bénéfice d'un Signifieur Central. Au lieu d'un tel flux unique (aussi vaste soit-il), il existe de multiples canaux dans lesquels des circuits spécialisés essaient, dans des pandémoniums parallèles, d'accomplir leurs diverses tâches, en créant des Versions Multiples au fur et à mesure qu'ils opèrent. La plupart de ces versions fragmentaires de « récit » jouent un rôle éphémère dans la modulation de l'activité du moment, mais certains se trouvent promus à un rôle fonctionnel nouveau, en se suivant rapidement, à travers l'activité d'une machine virtuelle existant dans le cerveau. Le caractère sériel de cette machine (son caractère « à la von Neumann ») n'est pas un trait d'organisation « câblée de façon innée » ; c'est plutôt le résultat d'une succession de coalitions de ces divers spécialistes.

Les spécialistes de base font partie de notre héritage animal. Ils n'ont pas été développés en vue d'accomplir des actions particulièrement humaines, telles que lire et écrire, mais des actions comme éviter des obstacles, échapper à des prédateurs, reconnaître des visages, saisir, envoyer, attraper des baies, et autres tâches essentielles. Ils sont souvent recrutés de façon opportuniste pour jouer de nouveaux rôles, pour lesquels leurs talents d'origine les rendent plus ou moins adaptés. Cela ne produit pas de tohu-bohu uniquement parce que les directions qui sont imposées à toute cette activité sont elles-mêmes le produit de l'organisation naturelle. Une partie de cette organisation est innée, et nous la partageons avec d'autres animaux. Mais elle est chez nous développée, et elle acquiert quelquefois une importance démesurée, par l'effet de microhabitudes de pensée qui sont développées dans l'individu, et qui sont le résultat en partie idiosyncrasique de l'exploration de soi et en partie le résultat des dons préconçus de la culture. Des milliers de mêmes, la plupart d'entre eux produits par le langage, mais aussi par des « images » non verbales et d'autres structures de données, prennent résidence dans un cerveau individuel, façonnant ses tendances et par là le transformant en un esprit.

Cette théorie est assez nouvelle pour être difficile à comprendre de prime abord, mais elle s'appuie sur des modèles développés par certains chercheurs en psychologie, en neurobiologie, en intelligence artificielle et en anthropologie, ainsi qu'en philosophie. Cet éclectisme imperturbable est souvent mis en question par les chercheurs appartenant aux domaines auxquels il emprunte ses éléments. En fréquentant ces différents domaines, je me suis habitué au manque de respect qu'expriment certains participants

pour leurs collègues appartenant aux autres disciplines. « Dan, pourquoi, demandent les chercheurs en intelligence artificielle, perds-tu ton temps à discuter avec ces types des neurosciences ? Ils font des allusions au “ traitement de l'information ” et ils se demandent où ce traitement s'effectue, et quels sont les neuro-transmetteurs qui entrent en jeu ; ils s'intéressent à tous ces faits barbants, mais ils n'ont pas la moindre idée des réquisits computationnels des fonctions cognitives supérieures. » « Pourquoi, demandent les chercheurs en neurosciences, perds-tu ton temps sur les fictions inventées par les chercheurs en intelligence artificielle ? Ils ne font qu'inventer n'importe quelle sorte de machinerie dont ils ont besoin, et tiennent des propos qui révèlent une ignorance impardonnable à propos du cerveau. » Les psychologues cognitifs, pendant ce temps, sont accusés de concocter des modèles qui n'ont *ni* plausibilité biologique *ni* pouvoirs computationnels démontrés ; les anthropologues ne reconnaîtraient pas un modèle s'ils en voyaient un, et les philosophes, comme nous le savons tous, ne font que se passer mutuellement le linge sale, en nous prévenant contre des confusions qu'ils ont eux-mêmes créées, dans un secteur dans lequel on ne trouve ni données ni théories empiriquement testables. Avec une telle bande d'idiots au travail sur le problème, il n'est pas étonnant que la conscience soit encore un mystère.

Toutes ces accusations sont correctes, et on peut en lancer encore bien d'autres, mais ce ne sont pas des idiots. Au contraire, la plupart des théoriciens auxquels j'ai emprunté des idées me frappent comme étant des individus intelligents – ce sont même des gens brillants, avec l'arrogance et l'impatience qui vont souvent avec le brio. Toutefois, leurs perspectives et leurs programmes sont limités : ils essayent de progresser sur des problèmes difficiles en empruntant les raccourcis qu'*eux* peuvent déceler, tout en déplorant ceux que prennent les autres. Personne ne peut être au clair sur tous les problèmes et sur tous les détails, moi-même y compris ; chacun doit piétiner, deviner et rester allusif sur de larges secteurs qui relèvent du problème examiné.

Par exemple, l'une des maladies professionnelles des neurosciences semble être la tendance à penser la conscience comme *le point terminal de la ligne*. (C'est comme d'oublier que le produit final visé par les pommiers n'est pas la pomme – mais plus de pommiers.) Bien entendu, ce n'est que récemment que les chercheurs en neurosciences se sont autorisés à réfléchir sur la cons-

cience ; et seul un petit nombre de théoriciens courageux ont commencé à dire officiellement ce qu'ils en pensaient. Comme le théoricien de la vision Bela Julesz l'a déclaré récemment, vous ne pouvez vraiment en parler que si vous avez les cheveux blancs – et un prix Nobel ! Voici, par exemple, une hypothèse qu'ont risquée récemment Francis Crick et Christof Koch :

« Nous avons suggéré que l'une des fonctions de la conscience est de présenter le résultat de diverses computations sous-jacentes et que cela implique un mécanisme attentionnel qui lie temporellement les neurones pertinents les uns aux autres en synchronisant leurs pointes dans des oscillations de 40 Hz » (Crick and Koch, 1990, p. 272).

Par conséquent, la fonction de la conscience serait de *présenter les résultats de computations sous-jacentes*. Mais à qui ? À la reine ? Crick et Koch ne vont pas plus loin et ne se posent pas la Question Difficile : *Et ensuite que se passe-t-il ?* (« Un miracle se produit-il ? ») À partir du moment où leur théorie a ramené une brebis dans ce qu'ils considèrent comme le cercle enchanté de la conscience, elle s'arrête. Elle n'aborde pas les problèmes que nous avons abordés dans les chapitres 4 à 7, par exemple ceux qui concernent la voie complexe qui va de la conscience (présumée) au comportement, y compris, en particulier, les comptes rendus introspectifs.

En revanche, les modèles de l'esprit offerts par la psychologie cognitive et l'IA ne souffrent presque jamais de ce défaut (voir, par exemple, Shallice, 1972, 1978 ; Johnson-Laird, 1983, 1988 ; Newell, 1990). Ils postulent en général un « espace de travail » ou une « mémoire de travail » qui remplace le Théâtre Cartésien ; ils montrent comment les résultats des computations qui y sont effectuées nourrissent d'autres computations qui guident le comportement, informent les comptes rendus verbaux, se doublent en retour récursivement de manière à fournir de nouvelles entrées à la mémoire de travail, et ainsi de suite. Mais ces modèles en général ne nous disent pas ou ni comment on peut localiser une mémoire de travail dans le cerveau, et s'occupent tellement du *travail* effectué dans cet espace de travail qu'il n'y a jamais un seul moment où il puisse y avoir du « jeu » – aucun signe de la *délectation* pour la phénoménologie qui semble être un trait si important de la conscience humaine.

Curieusement, donc, les chercheurs en neurosciences finissent

souvent par avoir l'air d'être dualistes, parce qu'une fois qu'ils ont « présenté » les choses dans la conscience, ils semblent passer la balle à l'Esprit, alors que les psychologues cognitifs finissent souvent comme des zombistes (des automatistes ?), parce qu'ils décrivent des structures ignorées des neuro-anatomistes, et parce que leurs théories visent à montrer que tout le travail peut s'effectuer sans avoir à sonner un Observateur Extérieur.

Les apparences sont trompeuses. Crick et Koch ne sont pas dualistes (même s'ils sont, apparemment, des matérialistes cartésiens), et les psychologues cognitifs n'ont pas *nié* l'existence de la conscience (même s'ils font de leur mieux, la plupart du temps, pour l'ignorer). Qui plus est, le fait que l'on approche ces phénomènes avec des œillères ne semble pas disqualifier l'une ou l'autre entreprise. Les chercheurs en neurosciences ont raison d'insister sur le fait que vous n'avez pas réellement un bon modèle de la conscience tant que vous n'aurez pas résolu le problème de savoir où elle peut se situer dans le cerveau. Mais les chercheurs en sciences cognitives (les gens de l'IA et les psychologues cognitifs, par exemple) ont raison d'insister sur le fait que vous n'avez pas réellement un bon modèle de la conscience tant que vous n'aurez pas résolu le problème de savoir quelles fonctions elle réalise et comment elle les réalise – mécaniquement, sans le bénéfice d'un Esprit. Comme dit Philip Johnson-Laird, « toute théorie scientifique de l'esprit doit le traiter comme un automate » (Johnson-Laird, 1983, p. 477). La perspective limitée qu'adopte chaque entreprise considérée en elle-même montre seulement que nous avons besoin d'une autre entreprise – celle dans laquelle nous sommes présentement engagés – qui essaie de rassembler autant de points forts de chacune que possible.

Comment s'orienter avec le croquis schématique

Ma tâche principale dans ce livre est philosophique : montrer comment une théorie de la conscience effectivement explicative *pourrait* être construite à partir de ces divers éléments, et non fournir – et confirmer – une telle théorie dans le détail. Mais ma théorie aurait été inconcevable (par moi, tout au moins) si elle n'avait pas emprunté beaucoup à des travaux empiriques dans différents domaines qui ont ouvert (pour moi en tout cas) de nouvelles façons de penser. (On peut trouver un recueil par-

ticulièrement riche de découvertes de ce genre et d'idées originales sur la conscience dans le livre de Marcel et Bisiach, 1988.) Nous appartenons à l'époque glorieuse de la recherche sur l'esprit. Toutes sortes de découvertes nouvelles, de nouveaux modèles, de résultats expérimentaux sont dans l'air — tout comme le sont d'ailleurs les prétendues « démonstrations » et rejets prématurés d'hypothèses. Présentement, la frontière de la recherche sur l'esprit est si largement ouverte qu'il n'y a presque pas de sagesse établie sur ce que peuvent être les bonnes questions et les bonnes méthodes. Quand il y a tant de fragments de théories et de spéculations mal étayées, il peut être bon de différer un peu nos exigences de démonstrations et de chercher plutôt des bases indépendantes mais également mal assurées qui tendraient à converger en faveur d'une unique hypothèse. Nous devrions cependant essayer de contenir notre enthousiasme. Il arrive qu'une fumée qui paraît suffisante pour entretenir un bon feu ne soit qu'un nuage de poussière venu d'un omnibus qui passe.

Dans *A Cognitive Theory of Consciousness* (1988), le psychologue Bernard Baars résume le « consensus grandissant » selon lequel la conscience est accomplie par une « société distribuée de spécialistes » qui est équipée d'une mémoire de travail, appelée *espace global de travail*, dont les contenus peuvent être diffusés au système « dans son ensemble » (p. 42). Comme il le note, tout un ensemble de théoriciens, en dépit d'énormes différences de perspective, de formation et d'aspirations, gravitent autour de cette vision commune de la position de la conscience dans le cerveau. C'est une version de ce consensus croissant que j'ai prudemment introduite, en ignorant certains traits et en en accentuant d'autres — à savoir les traits dont je pense qu'on les a négligés ou sous-estimés, et qui me semblent particulièrement cruciaux pour briser les mystères conceptuels qui demeurent encore.

Afin d'orienter ma théorie en relation à certaines des montagnes de travaux auxquels elle a fait des emprunts, revenons à mon petit croquis schématique, en examinant un seul thème à la fois, et en relevant les sources et les désaccords.

Il n'existe pas de « flux de conscience » unique, définitif, parce qu'il n'existe pas de Quartier Général central, pas de Théâtre Cartésien « où tout se rejoint » pour le bénéfice d'un Signifieur Central.

Tout le monde est d'accord pour dire qu'il n'existe pas pareil point de vue unique dans le cerveau, qui pourrait rappeler la glande pinéale de Descartes. Toutefois, on n'a pas reconnu quelles étaient les implications de cette idée, et on les néglige à l'occasion de façon frappante. Par exemple, des formulations peu prudentes du « problème du liage » dans la recherche actuelle en neurosciences présupposent souvent qu'il doit exister un espace représentationnel unique dans le cerveau (plus petit que le cerveau tout entier), où les résultats des diverses discriminations sont enregistrées les unes par les autres – mariant la bande-son avec le film, coloriant les formes, remplissant les parties laissées en blanc. Il existe certaines formulations prudentes du problème du liage (des problèmes de ce type) qui évitent cette erreur, mais on néglige souvent les points délicats.

Au lieu d'un tel flux unique (aussi vaste soit-il), il existe de multiples canaux dans lesquels des circuits spécialisés essaient, dans des Pandémoniums parallèles, d'accomplir leurs diverses tâches, en créant des Versions Multiples au fur et à mesure qu'ils opèrent. La plupart de ces versions fragmentaires de « récit » jouent un rôle éphémère dans la modulation de l'activité du moment....

En IA, l'importance des séquences de type narratif a été soulignée il y a longtemps par Roger Schank, d'abord dans son travail sur les *scripts* (1977, avec Abelson) et plus récemment (1991) dans son travail sur le rôle du récit dans la compréhension. À partir de perspectives très différentes, toujours à l'intérieur de l'IA, Patrick Hayes (1979), Marvin Minsky (1975), John Anderson (1983) et Erik Sandeval (1991) – entre autres – ont argumenté en faveur de l'importance de structures de données qui ne soient pas seulement des séquences de « clichés » (avec le problème afférent consistant à réidentifier des entités particulières dans des trames successives), mais sont plutôt construites de manière spécifique d'une façon ou d'une autre pour représenter des séquences temporelles et des types de séquences directement. En philosophie, Gareth Evans (1982) avait commencé à développer des idées parallèles avant sa mort précoce. En neurobiologie, on explore ces fragments narratifs sous le nom de *scénarios*, et d'autres séquences le sont dans l'approche de William Calvin (1987) avec sa Machine Darwinienne. Les anthropologues ont maintenu depuis longtemps que les mythes que transmet chaque culture à ses nouveaux membres jouent un rôle important dans

la formation de leur esprit (voir par exemple, Goody, 1977, et pour une application possible à l'IA, Dennett, 1991b), mais ils n'ont pas essayé d'en donner des modèles, computationnels ou neuro-anatomiques.

...mais certains se trouvent promus à des rôles fonctionnels nouveaux, en se suivant rapidement, à travers l'activité d'une machine virtuelle existant dans le cerveau. Le caractère sériel de cette machine (son caractère « à la von Neumann ») n'est pas un trait de conception « câblée de façon innée » ; c'est plutôt le résultat d'une succession de coalitions de ces divers spécialistes.

Beaucoup de gens ont remarqué le caractère relativement lent et laborieux de la démarche de l'activité consciente (par exemple Baars, 1988, p. 20) et on a suggéré depuis longtemps que cela pourrait être dû au fait que le cerveau n'est pas normalement conçu – câblé au départ – pour une telle activité. L'idée est dans l'air depuis plusieurs années : la conscience pourrait alors être l'activité d'une sorte de machine virtuelle sérielle implantée sur le matériel parallèle du cerveau. Le psychologue Stephen Kosslyn a offert une version de la machine virtuelle sérielle à une réunion de la Société pour la Philosophie et la Psychologie au début des années quatre-vingt, et j'ai esquissé diverses versions de cette idée à peu près à la même époque (par exemple Dennett, 1982b), mais on trouve une présentation d'une idée à peu près semblable – sans néanmoins que soit utilisé le terme de « machine virtuelle » – dans l'article pionnier du psychologue Paul Rozin, « The Evolution of Intelligence and Access to the Cognitive Unconscious », 1976. Un autre psychologue, Julian Jaynes, dans les spéculations audacieuses et originales que contient *The Origins of Consciousness and the Breakdown of the Bicameral Mind* (1976), a mis l'accent sur le fait que la conscience était une apparition très récente, produite par la culture sur une architecture fonctionnelle antérieure ; ce thème avait aussi été développé autrement par le neurobiologiste Harry Jerison (1973). L'architecture neuronale sous-jacente est loin d'être une *tabula rasa* ou un espace blanc à la naissance, selon cette conception ; elle est néanmoins un véhicule dans lequel les structures se construisent en fonction des interactions du cerveau avec le monde. Et ce sont ces structures déjà construites, plus que les structures innées, que l'on doit évoquer pour expliquer le fonctionnement cognitif.

Les spécialistes de base font partie de notre héritage animal. Ils n'ont pas été développés en vue d'accomplir des actions particulièrement humaines, telles que lire et écrire, mais des actions comme éviter des obstacles, échapper à des prédateurs, reconnaître des visages, saisir, envoyer, attraper des baies, et autres tâches essentielles...

Ces hordes de spécialistes sont attestées par des théories très différentes, mais leur taille, leur rôle et leur organisation font l'objet de débats animés. (Pour un résumé rapide et très utile voir Allport, 1989, p. 643-647.) Les neuro-anatomistes qui étudient le cerveau des animaux, depuis les seiches et des calamars jusqu'aux chats et aux singes, ont identifié beaucoup de variétés de circuits câblés de façon innée et remarquablement conçus pour accomplir certaines tâches. Les biologistes parlent de Mécanismes Innés de Déclenchement (*Innate Releasing Mechanisms*, IRM) et de Trames d'Action Fixes (*Fixed Action Patterns* (FAP)), qui peuvent se trouver assemblés, et dans une lettre récente qu'il m'a adressée, le neuropsychologue Lynn Waterhouse décrit très bien les esprits des animaux comme composés de « matelas d'IRM-FAP ». Ce sont précisément ces esprits animaux matelassés problématiquement dont Rozin (parmi d'autres) présuppose qu'ils sont à la base de l'évolution d'esprits à fonctions plus générales. Le psychologue de la perception V.S. Ramachandran (1991) note qu'« il existe un avantage effectif qui apparaît dans des systèmes multiples : il vous achète de la tolérance pour les images bruitées comme celles que vous rencontrez dans le monde réel. Mon analogie favorite pour illustrer certaines de ces idées est que c'est un peu comme deux hommes ivres : aucun d'entre eux ne peut marcher sans être soutenu mais en s'appuyant l'un sur l'autre ils parviennent à tituber vers leur objectif ».

Le neuropsychologue Michael Gazzaniga a attiré l'attention sur une masse de données provenant de déficits neurologiques (y compris le cas fameux, mais souvent décrit de façon erronée, des patients au cerveau divisé) qui étayaient une conception de l'esprit comme coalition ou comme amas d'agences semi-indépendantes (Gazzaniga et Ledoux, 1978 ; Gazzaniga, 1985). Venant d'un camp différent, le philosophe de la psychologie Jerry Fodor (1983) a soutenu que de larges parties de l'esprit humain sont composées de *modules* : des systèmes câblés de façon innée, à fonctions spécifiques, « encapsulés », chargés d'analyser des entrées

(et d'engendrer des sorties – bien qu'il n'ait pas grand-chose à dire sur ce dernier point).

Fodor se concentre sur les modules qui seraient spécifiques à l'esprit humain – des modules destinés à acquérir le langage et à analyser les phrases, en particulier – et puisqu'il ignore dans l'ensemble le problème de savoir ce que pourraient être leurs ancêtres probables dans les esprits des animaux inférieurs, il crée l'impression improbable que l'évolution a conçu des mécanismes tout neufs spécifiques à chaque espèce pour le langage habillés de neuf, pour ainsi dire. Cette image des modules comme don miraculeux qu'aurait fait Dame Nature à l'*Homo sapiens* est encouragée par la conception ultra-intellectualiste qu'a Fodor de la manière dont les modules sont attachés au reste de l'esprit. Selon lui, ils n'accomplissent pas des tâches complètes dans l'économie de l'esprit (comme contrôler une coordination main-œil pour attraper quelque chose) mais s'arrêtent abruptement sur une limite interne, une ligne dans l'esprit qu'ils ne peuvent franchir. Il y a une aire centrale de « fixation de la croyance », soutient Fodor, dans laquelle les modules déposent, tels des esclaves, leurs biens, en les transformant en processus non modulaires (« globaux, isotropes »).

Les modules de Fodor sont un rêve de bureaucrate : les descriptions de leur travail sont gravées dans le roc ; ils ne peuvent être recrutés pour jouer des rôles nouveaux ou multiples, et ils sont « cognitivement impénétrables » – ce qui veut dire que leurs activités ne peuvent pas être modulées, ou même interrompues, par des changements dans les états informationnels « globaux » du reste du système. Selon Fodor, toutes les activités de la cognition qui impliquent effectivement la pensée sont non modulaires. Se représenter ce qu'on doit faire ensuite, raisonner sur des situations hypothétiques, restructurer ses données de façon créatrice, réviser sa conception du monde, toutes ces activités sont accomplies par un dispositif central mystérieux. De plus, Fodor soutient (avec un curieux sentiment de satisfaction) qu'aucune branche des sciences cognitives – y compris la philosophie – n'a la moindre idée de la manière dont ce dispositif central pourrait fonctionner !

« On en sait beaucoup sur les transformations des représentations qui servent à mettre l'information sous une forme qui soit appropriée pour le traitement central ; on ne sait pratiquement rien de ce qui se passe après que l'information est passée par là. Le fantôme

a été repoussé plus loin dans la machine, mais on ne l'a pas exorcisé » (Fodor 1983, p. 127).

En accordant au dispositif central un tel travail et un pouvoir non modulaire si vaste pour l'effectuer, Fodor transforme ses modules en agents très peu plausibles ; leur existence n'a de sens qu'en compagnie d'un agent en chef doué d'une autorité fantastique (Dennett, 1984b). Puisque l'une des thèses principales de Fodor, quand il décrit les modules, consiste à opposer leur mécanicité finie, compréhensible, stupide, aux pouvoirs illimités et inexplicables du centre non modulaire, les théoriciens qui seraient sans cela réceptifs presque à la majorité de cette caractérisation des modules ont vu dans ses modules des fictions crypto-cartésiennes.

Ces mêmes théoriciens, ou du moins beaucoup d'entre eux, ont oscillé entre la tiédeur et l'hostilité face aux *agents* de Marvin Minsky, qui forment *la société de l'esprit* (1985). Les agents de Minsky sont des homoncles qui prennent toutes les tailles, des spécialistes géants doués de talents à peu près aussi élaborés que ceux des modules fodoriens, jusqu'aux agents qui ont la taille des mêmes (les polymèmes, les micromèmes, les agents censeurs, les agents supprimeurs, et bien d'autres). Cela a l'air trop facile, pensent les sceptiques. Chaque fois qu'il y a une tâche, postulez une bande d'agents de même taille que la tâche pour l'accomplir : c'est une démarche théorique qui a toutes les vertus du vol par rapport au travail honnête, pour paraphraser une expression fameuse de Russell.

Les homoncles – démons, agents – sont les pièces de monnaie du royaume de l'intelligence artificielle, et de l'informatique en général. Quiconque éprouve des démangeaisons sceptiques quand on mentionne seulement des homoncles ne comprend tout simplement pas combien ce concept peut être neutre, et combien il est applicable dans de nombreux cas. Postuler une bande d'homoncles serait bien entendu un geste aussi vide que l'imaginer le sceptique, si n'intervenait pas le fait suivant : dans les théories des homoncles, le contenu sérieux tient aux affirmations portant sur les modes d'interactions des homoncles postulés, sur leurs modes d'interaction, développement, et sur la manière dont ils forment des coalitions et des hiérarchies, et ainsi de suite. Et ici les théories peuvent en fait être très différentes les unes des autres. Les théories bureaucratiques, comme nous l'avons vu dans

le chapitre 7, organisent les homoncules en hiérarchies pré-construites. Il n'existe pas d'homoncules subventionnés ou créateurs de désordres, et la compétition entre les homoncules est étroitement conçue comme une partie de base-ball de championnat. Les théories du Pandémonium, en revanche, postulent une grande quantité de duplications d'efforts, des mouvements en pure perte, de l'interférence, des périodes de chaos, et des saisonniers sans tâches fixées à l'avance. Appeler homoncules (ou démons ou agents) les unités qui font partie de ces différentes théories est à peine plus sensé que de les appeler simplement... des unités. Ce sont seulement des unités pourvues de compétences particulières définies, et toute théorie, de celles qui sont le plus rigoureusement neuro-anatomiques à celles qui sont le plus abstraitement artificielles, postule des unités de ce genre et théorise ensuite sur la manière dont les fonctions plus larges peuvent être réalisées par des organisations d'unités accomplissant des fonctions plus réduites. En fait, *toutes* les variétés de fonctionnalisme peuvent être considérées comme du fonctionnalisme « homonculaire » à un grain de description ou à un autre.

J'ai été amusé de relever une sorte d'euphémisme qui a récemment eu les faveurs des chercheurs en neurosciences. Les neuro-anatomistes ont fait récemment des avancées énormes dans la cartographie du cortex, qui se trouve être délicatement organisée en colonnes spécialisées de neurones interactifs (le neurophysiologiste Vernon Mountcastle, 1978, les appelle « modules unités »), qui sont constitués en organisations plus larges telles que les « cartes rétinotopiques » (dans lesquelles la trame spatiale d'excitation sur les rétines des yeux est préservée), qui en retour jouent des rôles – que nous ne comprenons pas toujours – dans des organisations de neurones ultérieures. Jadis, les chercheurs en neurosciences parlaient de ce que ces divers nappes ou groupes de neurones dans le cortex *signalaient* ; ils considéraient que ces unités étaient des homoncules dont le « boulot » était toujours « d'envoyer un message doté d'un contenu particulier ». Les progrès récents de la théorie neuro-anatomique ont suggéré que ces nappes de neurones accomplissaient des fonctions encore plus complexes et variées, en sorte qu'on voit à présent qu'il est trompeur de les considérer comme signalant simplement ceci ou cela. Comment, par conséquent, pourrions-nous exprimer les découvertes péniblement acquises sur les conditions spécifiques sous lesquelles ces nappes sont actives ? Nous disons que ce tract

« est attentif » à la couleur, alors qu'un autre « est attentif » au lieu ou au mouvement. Mais cet usage ne revient pas à un anthropomorphisme ridicule ou à un « paralogisme de l'homomconcule » du type de celui que l'on rencontre partout en IA ! Dieu merci, ce n'est pas le cas : ce n'est qu'une façon habile pour les chercheurs sobres de s'exprimer de façon suggestive sans vouloir faire de l'effet et de parler de la compétence de ces nappes de neurones ! On peut les manger tous deux à la même sauce.

Les agents de Minsky ont ceci de particulier que, à la différence de la plupart des autres sortes d'homoncles postulés, ils ont des histoires et des généalogies. Leur existence n'est pas seulement postulée : ils doivent se développer à partir de quelque chose dont l'existence antérieure n'était pas entièrement mystérieuse, et Minsky a fait de nombreuses suggestions sur la manière dont ce genre de développement a pu se produire. S'il garde encore une attitude neutre qui peut être agaçante quant à la nature exacte des neurones agents et quant à leur localisation dans le cerveau, c'est simplement parce qu'il a voulu explorer les réquisits généraux du développement des fonctions sans spécifier trop leur nature. Comme il le dit, en décrivant sa théorie antérieure des « cadres » (dont la Société de l'Esprit est le descendant) : « Si la théorie avait été plus vague, elle aurait été ignorée, mais si elle avait été décrite avec plus de détails, les autres savants l'auraient " testée ", plutôt que de chercher à apporter la contribution de leur propre idée » (1985, p. 259). Certains savants restent froids face à ce genre d'apologie. Ils ne s'intéressent qu'aux théories qui produisent des prédictions qu'on peut tester tout de suite. Cela serait de bonne politique, à cette nuance près que toutes les théories testables inventées jusqu'à présent ont été réfutées, et il est vain de penser que les percées requises pour composer de nouvelles théories testables sortiront comme par enchantement sans les efforts d'imagination comparables à ceux que Minsky se permet. (J'ai fait de même, bien entendu.)

Revenons à notre schéma récapitulatif :

Ils [les spécialistes] sont souvent recrutés de façon opportuniste pour jouer de nouveaux rôles, pour lesquels leurs talents d'origine les rendent plus ou moins adaptés. Le résultat n'est pas du tohu-bohu seulement parce que les directions qui sont imposées à toute cette activité sont elles-mêmes le produit de l'organisation naturelle. Une partie de cette organisation est innée, et nous la partageons avec d'autres animaux. Mais elle est chez nous développée,

et elle acquiert quelquefois une importance démesurée, par l'effet de microhabitudes de pensée qui sont développées dans l'individu, et qui sont les résultats en partie idiosyncrasiques de l'exploration de soi et en partie des dons préconçus de la culture. Des milliers de mêmes, la plupart d'entre eux produits par le langage, mais aussi par des « images » non verbales et d'autres structures de données, prennent résidence dans un cerveau individuel, façonnant ses tendances et par là le transformant en un esprit.

Dans cette partie de ma théorie j'ai pris volontairement garde de ne pas m'engager sur diverses questions importantes : comment ces homoncules entrent-ils effectivement en interaction pour accomplir quoi que ce soit ? En quoi consistent les transactions de traitement de l'information effectives, et quelles raisons avons-nous de penser qu'elles pourraient « marcher » ? Selon le schéma que j'ai donné, la suite des événements est déterminée (de façons que je n'ai qu'esquissées) par des « habitudes », et mis à part diverses thèses négatives proposées au chapitre sur ce qui *ne se passe pas*, je n'ai pas été très précis quant à la structure des processus par lesquels divers éléments provenant des Versions Multiples sont perpétués, certains d'entre eux engendrant finalement l'hétérophénoménologie par l'intermédiaire d'un sondage ou d'un autre. Pour voir de quoi il est question, et ce que pourraient être les diverses réponses, nous devrions considérer brièvement certains modèles explicites de pensée séquentielle.

Et que se passe-t-il ensuite ?

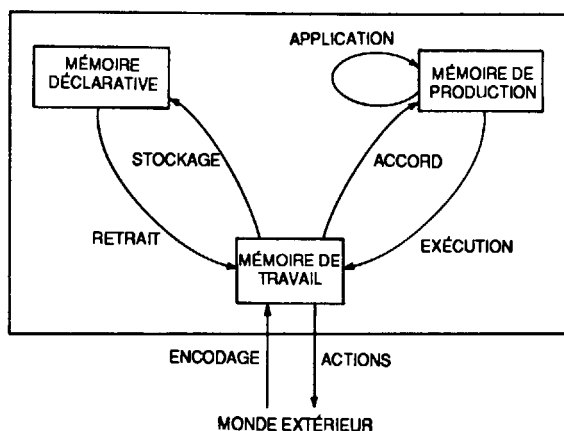
Dans le chapitre 6, nous avons vu comment l'architecture de von Neumann était une quintessence du processus sériel de calcul délibéré. Turing et von Neumann ont isolé une sorte particulière de courant qui peut circuler à travers le flux de la conscience, et ils l'ont radicalement idéalisé pour les besoins de la mécanisation. Le célèbre goulot de von Neumann consiste en un registre unique de résultats et un registre unique d'instructions. Les programmes ne sont que des listes ordonnées d'instructions tirées d'un petit ensemble de notions primitives que la machine est câblée au départ pour exécuter. Un processus fixe, le cycle recherche-exécution, tire les instructions de la queue dans la mémoire, une seule à la fois, prenant toujours l'instruction qui

suit dans la liste, à moins que l'instruction précédente ne se raccorde à une autre partie de la liste.

Quand les modélisateurs en IA se sont mis à implémenter des modèles plus réalistes des opérations cognitives sur cette base, ils ont révisé tout cela. Ils ont augmenté la taille du goulot de von Neumann qui était bien trop étroit et l'ont transformé en un « espace de travail » ou une « mémoire de travail » plus compréhensifs. Ils ont aussi conçu des opérations plus sophistiquées chargées de jouer le rôle de notions psychologiques primitives et remplacé le cycle d'instructions rechercher-exécuter de la machine de von Neumann par des procédures plus flexibles d'appel et d'exécution des instructions. L'espace de travail devient, dans certains cas, un « tableau noir » (Reddy et al., 1973 ; Hayes-Roth, 1985), sur lequel divers démons pouvaient écrire des messages que tous les autres puissent lire, ce qui provoqua en retour une autre vague d'écriture et de lecture. L'architecture de von Neumann, avec son cycle rigide d'instructions, était encore là, dans l'arrière-plan, réalisant l'implémentation, mais elle ne jouait pas de rôle effectif dans le modèle. Dans le modèle, ce qui devait se passer ensuite était gouverné par les résultats des vagues d'écriture et de lecture de messages sur le tableau noir qui entraient en compétition. Une autre espèce voisine de descendants de l'architecture de von Neumann sont les divers *systèmes de production* (Newell, 1973) qui sous-tendent des modèles comme celui de modèle ACT * John Anderson (1983) (prononcer act – étoile) et le modèle Soar de Rosebloom, Laird et Newell (1987, voir aussi Newell, 1990).

On peut se faire une idée de l'architecture qui sous-tend un système de production sur ce diagramme simple de ACT * (figure 8.1).

La mémoire de travail est l'endroit où les choses se passent. Toutes les actions de base s'appellent des *productions*. À la base, les productions ne sont que des mécanismes de reconnaissance de trames qui sont ajustées de manière à se déclencher chaque fois qu'elles détectent *leur* trame particulière. Cela veut dire que ce sont des opérateurs SI-ALORS qui se promènent en regardant les contenus présents de la mémoire de travail, en attendant que leurs conditions SI soient satisfaites, de façon à ce qu'ils puissent ALORS faire leur boulot, quel qu'il soit (dans un système classique de production, cela dépose un nouvel élément de donnée dans la mémoire de travail, pour l'usage ultérieur des productions).



De *The Architecture of Cognition*, par John A. Anderson, Cambridge, Mass. Harvard University Press, © 1983, par le Président et les Fellows de Harvard College.

Figure 8.1

Tous les ordinateurs ont des primitives SI-ALORS, « organes sensoriels » qui leur permettent de réagir différemment à des données entrantes ou retirées de la mémoire de travail. Cette capacité de *branchement conditionnel* est un ingrédient initial pour le pouvoir de l'ordinateur, quelle que soit la nature de l'architecture. Les SI-ALORS initiaux étaient les instructions simples, nettes, des états de la machine de Turing : SI vous voyez un zéro, ALORS remplacez-le par un un, déplacez un espace vers la gauche, et passez en l'état *n*. Comparez avec des instructions simples du type de celles que vous pourriez donner à un système sensoriel humain bien entraîné et expérimenté : SI vous voyez quelque chose qui vous paraît familier, ET si une analyse supplémentaire ne résout pas le problème OU si vous avez des doutes résiduels, ALORS sonnez l'alarme. Pouvons-nous construire de la gestion sophistiquée de données à partir de SI-ALORS mécaniques de ce type ? Les productions sont des senseurs de niveau intermédiaire à partir desquels on peut construire des organes des sens plus complexes, puis des architectures globales de cognition. Les productions peuvent avoir des conditions SI complexes et mal délimitées : les trames qu'elles « reconnaissent » n'ont pas à être aussi simples que les codes-barres que reconnaissent les caisses enregistreuses, mais plutôt semblables aux trames que reconnaît le système sensoriel (voir la discussion dans Anderson, 1983, p. 35-44). Et à la différence des SI-ALORS d'une machine de Turing, qui est toujours dans un seul état à la fois (toujours en train de

tester un SI-ALORS de cet ensemble, avant de passer à la donnée suivante), les SI-ALORS d'un système de production attendent en masse en parallèle (simulé), en sorte qu'à un « instant » quelconque plus d'une production peut voir sa condition satisfaite et être prête à l'action.

C'est là que les choses deviennent intéressantes : comment un tel système se comporte-t-il quand il s'agit de *résoudre des conflits* ? Quand plus d'une production est satisfaite, il y a toujours une chance pour que deux (ou plus) tirent dans des directions incompatibles. Les systèmes parallèles peuvent tolérer une grande quantité de conflits entre les tâches, mais dans un système qui doit réussir dans le monde tout ne peut pas se passer en même temps ; il arrive qu'il faille laisser tomber quelque chose. La façon dont sont résolus les conflits est une différence fondamentale entre divers systèmes. En fait, puisque la plupart, sinon tous les détails qui ont un intérêt psychologique et biologique, se situent dans les différences repérées à ce niveau, il vaut mieux considérer l'architecture du système de production comme un véhicule sous-jacent dans lequel on peut construire des modèles. Mais tous les systèmes de production partent de quelques principes de base qui servent de relais à notre esquisse théorique : ils ont un espace de travail où l'action se produit, et où de nombreuses productions (= des démons) essaient de faire leur propre truc en même temps, et ils ont une mémoire plus ou moins inerte où de l'information accumulée et innée se trouve stockée. Puisque tout ce qui est « connu » par un tel système n'est pas disponible en même temps, le problème de Platon – attraper le bon oiseau au bon moment – est la tâche logistique majeure à laquelle ils font face. Et, chose plus importante de notre point de vue ici, les théoriciens ont *effectivement construit* des mécanismes destinés à répondre à la Question Difficile : *que se passe-t-il ensuite ?*

Par exemple, dans ACT *, on trouve cinq principes de résolution de conflits :

1. *Degré de correspondance* : si une condition SI de production est d'une manière ou d'une autre mieux satisfaite qu'une autre, elle a la priorité.

2. *Force de production* : les productions qui ont récemment eu du succès se voient associer une « force » plus grande, qui leur donne la priorité sur les productions plus faibles.

3. *Caractère réfractaire des données* : la même production ne peut pas correspondre aux mêmes données plus d'une fois (pour empêcher les boucles infinies et les excitations similaires moins drastiques).

4. *Spécificité* : quand deux productions correspondent aux mêmes données, la production qui a la clause si la plus spécifique gagne.

5. *Dominance du but* : parmi les entités que les productions déposent dans la mémoire de travail se trouvent les buts. Il ne peut y avoir qu'un seul but actif à un moment donné dans la mémoire de travail d'ACT *, et toute production dont la sortie correspond au but actif a la priorité.

Ce sont là tous les principes plausibles de résolution de conflits, qui ont un sens à la fois du point de vue psychologique et du point de vue téléologique (pour une discussion détaillée, voir Anderson, 1983, ch. 4). Mais peut-être sont-ils trop évidents. Cela veut dire qu'Anderson a lui-même conçu avec sagesse le système de résolutions de conflits d'ACT *, en exploitant sa connaissance des types spécifiques de problèmes qui surviennent dans des circonstances de résolution de conflit, et sa connaissance des manières effectives de les traiter. Il a câblé pour l'essentiel dans le système lui-même cette connaissance sophistiquée, en en faisant un don inné de l'évolution. On peut faire une comparaison intéressante avec l'architecture Soar de Rosenbloom, Laird et Newell (1987). Elle aussi, comme toute architecture parallèle, rencontre des *impasses* – des occasions où il faut résoudre un conflit parce que se « déclenchent » soit des productions contradictoires soit aucune production – mais elle les traite comme un avantage, pas comme un problème. Les impasses sont les principales occasions que le système a de se construire. Les conflits ne sont pas automatiquement traités par un ensemble préconçu et fixe de principes de résolution de conflit (un policier homoncule autoritaire chargé de la circulation et déjà en place) ; ils sont plutôt traités de façon *non* automatique. Une impasse crée un *nouvel* « espace de problèmes » (une sorte d'espace de travail topique) dans lequel le problème à résoudre est précisément l'impasse. Cela peut engendrer encore un autre métaméta-espace de problème de trafic, et ainsi de suite – *potentiellement* à l'infini. Mais en pratique (au moins dans les domaines modélisés jusqu'à présent), après avoir enregistré les espaces de problèmes sur plusieurs niveaux, le

problème en haut de la pile se trouve résolu, ce qui résout rapidement le problème au niveau inférieur suivant, et ainsi de suite, dissolvant ainsi l'énorme prolifération des espaces après une exploration non triviale dans l'espace logique des possibilités. L'effet sur le système, qui plus est, est de « regrouper » les découvertes acquises haut la main en nouvelles productions en sorte que quand de nouveaux problèmes surgissent dans le futur, il y a une nouvelle production inventée disponible pour les résoudre rapidement, qui devient ainsi un problème trivial déjà résolu dans le passé.

Je mentionne rapidement ces détails non pour défendre les mérites qui feraient que Soar l'emporterait sur ACT *, mais pour donner une idée du type de problèmes que l'on peut, avec sérieux, explorer en utilisant des modèles ainsi composés. J'ai l'intuition que, pour diverses raisons qui n'ont pas besoin de nous concerner ici, le véhicule sous-jacent des systèmes de production est *encore* trop idéalisé et trop simplifié dans ses contraintes. Mais la trajectoire qui va de la machine de von Neumann vers les systèmes de production indique encore d'autres architectures, dont la structure ressemble encore plus à celle du cerveau, et la meilleure façon d'explorer leurs forces et leurs limitations est de les construire et de les faire marcher. C'est ainsi que l'on pourra transformer des théories impressionnistes et vagues comme la mienne en modèles honnêtes dont on puisse empiriquement tester les détails.

Quand vous réunissez les diverses thèses sur les mécanismes de la conscience que j'ai avancées dans les quatre derniers chapitres et que vous essayez de les juxtaposer à des modèles de systèmes cognitifs comme ceux dont il vient d'être question, de nombreuses questions surgissent, mais je ne vais pas essayer ici de répondre à toutes. Puisque je laisse toutes ces questions sans réponse, mon ébauche en reste là : c'est une esquisse qui pourrait correspondre de façon lâche à toute une famille de théories différentes. En cette occasion, je ne peux aller plus loin, parce que les problèmes philosophiques de la conscience portent sur la question de savoir si *une théorie quelconque* de ce type peut expliquer la conscience, en sorte qu'il serait prématuré d'épingler nos espoirs sur une version trop particulière qui pourrait se révéler sérieusement fautive. (Dans l'appendice B, j'ajouterai quelques éléments empiriques, pour ceux qui veulent qu'une théorie ait des implications qu'on puisse tester dès le départ.)

Les théories philosophiques ne sont pas les seules à avoir besoin d'être mises en bonne et due forme en modélisant à ce niveau : en neurosciences, les théories sont sur le même bateau. Par exemple, la théorie complexe de Gerald Edelman (1990) sur les circuits de « ré-entrée » dans le cerveau peut faire les discriminations, construire les structures de mémoire, coordonner les suites d'étapes de résolution de problèmes, et en général exécuter les activités d'un esprit humain, mais en dépit de la richesse de ses détails anatomiques, et des assertions enthousiastes et souvent plausibles d'Edelman lui-même, nous ne savons pas ce que ses ré-entrées peuvent faire – nous ne savons pas si les ré-entrées sont la façon *correcte* de concevoir la neuro-anatomie fonctionnelle – tant qu'elles ne sont pas incorporées dans une architecture cognitive globale qui ait le même grain de précision que ACT* ou Soar et tant qu'elle ne suit pas leur voie².

Si l'on adopte une modélisation plus fine, il restera un travail non achevé : montrer comment les productions elles-mêmes (ou les démons reconnaisseurs de trames, quel que soit le nom que nous leurs donnions) peuvent être implémentées dans le cerveau. Baars (1988) appelle ses spécialistes des « briques » avec lesquelles on peut construire, et il choisit de laisser les détails plus profonds des constructions avec les briques pour une autre fois ou pour une autre discipline, mais, comme l'ont noté beaucoup de critiques, il est tentant de supposer que les spécialistes eux-mêmes, à plusieurs niveaux d'agrégation, devraient être modélisés comme étant faits de fabriques *connexionnistes* d'un type ou d'un autre.

Le connexionnisme (ou PPD, *Parallel Distributed Processing*, Traitement Parallèle Distribué) est un développement assez récent en IA qui promet de rapprocher la modélisation cognitive de la modélisation neuronale, puisque les éléments qui constituent ses briques sont des nœuds dans des réseaux parallèles connectés de différentes façons qui ressemblent *assez* à des réseaux de neurones dans le cerveau. Comparer l'IA connexionniste à la « bonne vieille intelligence artificielle » (Haugeland, 1985) et à divers projets de modélisation en neurosciences est devenu une industrie universitaire majeure (voir par exemple Graubard, 1988, Bechtel et Abrahamson, 1991, Ramsey, Stich et Rumelhart, 1991). Ce n'est pas surprenant, parce que le connexionniste fraie les premiers chemins d'unification qui soient plausibles dans l'énorme *terra incognita* qui sépare les sciences de l'esprit et les sciences du cerveau. Mais presque aucune des controverses qui entourent le

« traitement correct du connexionnisme » (Smolensky, 1988) n'affecte nos projets ici. *Bien entendu*, il faudra qu'il existe un niveau (ou des niveaux) théorique(s) qui ai(en)t à peu près la même précision que les modèles connexionnistes, pour servir de médiation entre des niveaux théoriques plus explicitement neuro-anatomiques et des niveaux théoriques plus explicitement psychologiques ou cognitifs. Le problème est de savoir quelles seront celles des idées connexionnistes particulières qui feront partie de cette solution et celles qui seront hors course. Tant que cette question n'est pas réglée, les chercheurs tendent à utiliser l'arène du débat connexionniste comme un amplificateur pour leurs slogans favoris, et bien que j'aie tout autant envie de prendre parti dans ces débats que n'importe qui (Dennett, 1987b, 1988b, 1989, 1990c, 1991b,c,d) je me mordrai la langue en cette occasion, et je me consacrerai à ma tâche principale, qui consistera à examiner en quoi une théorie de la *conscience* pourrait émerger de tout cela quand la poussière aura commencé à recouvrir ces débats, quelle que soit la façon dont cela se fera.

Notons ce qui s'est produit au cours de la progression de l'architecture de von Neumann aux architectures virtuelles en tant que systèmes de production et aux systèmes connexionnistes (à un niveau plus fin de description). Il s'est produit ce que l'on pourrait appeler un déplacement dans l'équilibre du pouvoir. Des programmes fixes, préconstruits, avançant sur des rails avec peu de points de branchement selon les données, ont été remplacés par des systèmes flexibles – en fait volatiles – dont le comportement ultérieur est bien plus fonction d'interactions complexes entre ce que le système rencontre sur le moment et ce qu'il a rencontré dans le passé. Comme Newell, Rosenbloom et Laird (1989) le disent, « ainsi le problème pour l'ordinateur standard est celui de savoir comment il doit être interrompu, alors que le problème pour Soar et ACT* (et sans doute aussi pour la cognition humaine) est de savoir comment il peut rester fixé à sa tâche » (p. 119).

Vue la masse d'encre qu'on a répandue sur ce problème théorique, il est important de mettre l'accent sur le fait qu'il y a un déplacement dans l'équilibre du pouvoir et non vers un mode d'opération « qualitativement différent ». Au cœur du système de reconnaissance de formes le plus volatile (qu'il soit « connexionniste » ou pas), on trouve un moteur à la von Neumann qui avance péniblement, calculant sa fonction calculable. Depuis la

naissance des ordinateurs, les critiques de l'intelligence artificielle n'ont cessé de dénoncer la rigidité, le caractère mécanique, le caractère *programmé* des ordinateurs, et ses défenseurs ont sans relâche souligné que c'était en fait une affaire de degré de complexité – que des systèmes organiques indéfiniment non rigides, vagues, holistiques, pouvaient être créés sur un ordinateur. Au fur et à mesure que l'IA s'est développée, des systèmes de ce genre sont apparus, si bien qu'aujourd'hui, les critiques ont à décider s'ils doivent pêcher ou désarmer. S'ils déclaraient que les systèmes connexionnistes – par exemple – sont les sortes de choses qu'ils avaient toujours pensé qu'étaient les esprits humains, ou s'ils levaient la mise et insistaient sur le fait que même un système connexionniste n'est pas suffisamment « holistique », ou suffisamment « intuitif » à leurs yeux, ou (placez ici votre slogan favori). Deux des critiques les plus célèbres de l'IA, les philosophes de Berkeley John Searle et Hubert Dreyfus, se divisent sur ce sujet : Dreyfus a fait acte d'allégeance au connexionnisme (Dreyfus et Dreyfus, 1988), alors que Searle a levé la mise, en insistant sur le fait qu'aucun ordinateur connexionniste ne pourrait manifester de véritables traits mentaux (1990a, 1990b).

Les sceptiques « par principe » peuvent bien faire retraite, mais les unificateurs rencontreront encore d'énormes problèmes. Le plus important, à mon sens, est un problème qui a un impact direct sur notre théorie de la conscience. Le consensus en sciences cognitives, que l'on pourrait illustrer par des douzaines de diagrammes comme la figure 8.1, est que *là* nous avons la mémoire à long terme (le colombier de Platon) et *ici* nous avons l'espace de travail ou la mémoire de travail, où la pensée se produit, en fait ³. Et pourtant, il n'y a pas deux endroits dans le cerveau où l'on pourrait loger ces deux ressources séparées. Le seul endroit dans le cerveau qui soit un réceptacle plausible pour l'une ou l'autre de ces fonctions séparées est l'ensemble du cortex – nous n'avons pas deux lieux l'un à côté de l'autre, mais un seul lieu étendu. Comme le dit Baars, qui résume ici le consensus grandissant, il existe un espace de travail *global*. Il est global non seulement au sens fonctionnel (en gros, c'est un « lieu » où à peu près tout peut entrer en contact avec tout le reste), mais aussi au sens anatomique (il est distribué sur tout le cortex et implique sans doute d'autres régions du cerveau). Cela veut dire, par conséquent, que l'espace de travail doit disposer des mêmes nappes

de neurones et des mêmes réseaux que ceux qui jouent apparemment un rôle majeur dans la mémoire de travail : le « stockage » des changements de conception qui sont provoqués par l'exploration individuelle.

Supposez que vous appreniez comment faire du pain avec de la farine de maïs ou que vous appreniez ce que « phénotypique » veut dire. D'une façon ou d'une autre, le cortex doit être un véhicule dans lequel des structures de connexions stables peuvent ancrer ces modifications de conceptions de façon tout à fait permanente dans le cerveau avec lequel vous êtes né. Supposez que l'on vous rappelle soudain votre rendez-vous avec le dentiste, et que cela dissipe tout le plaisir que vous tiriez de la musique que vous étiez en train d'écouter. D'une certaine manière, le cortex doit être un véhicule dans lequel des structures de connexions *instables* peuvent rapidement altérer ces contenus transitoires de l'« espace » tout entier – sans, bien entendu, effacer la mémoire à long terme dans le processus. Comment ces deux sortes de « représentation » peuvent-elles coexister dans le même véhicule au même moment ? Dans les modèles purement cognitifs, ces fonctions peuvent figurer dans des boîtes séparées sur un diagramme, mais quand nous avons à les superposer sur une fabrique unique de tissus neuronaux, le problème simple de l'empaquetage est le moindre de nos soucis.

On peut supposer que deux systèmes en réseau s'interpénètrent (tout comme le système téléphonique et le système des autoroutes couvrent le continent) – ce n'est pas le problème. Le problème plus profond se pose juste au-dessous de la surface, dans une hypothèse que nous avons faite. Les démons spécialistes individuels, avons-nous supposé, recrutent en quelque sorte les autres au sein d'une entreprise plus vaste. S'il s'agissait simplement d'appeler ces nouvelles recrues à exercer leurs talents de *spécialistes* au service d'une cause commune, nous aurions déjà des modèles de ce genre de processus – comme ACT *, Soar et l'Espace de Travail Global de Baars – avec des descriptions plus ou moins plausibles de leur structure de détail. Mais que se passe-t-il si les spécialistes sont aussi recrutés quelquefois *en tant que généralistes*, pour contribuer à des fonctions dans lesquelles leurs talents de spécialistes ne jouent pas de rôle discernable ? L'idée est tentante, pour diverses raisons (voir, par exemple, Kinsbourne et Hicks, 1978), mais, pour autant que je sache, nous n'avons

toujours pas de modèles computationnels de la façon dont ces éléments à double fonction peuvent opérer.

La difficulté est la suivante : on suppose couramment que les spécialistes dans le cerveau doivent tenir en quelque manière leur identité fonctionnelle de leur position effective dans un réseau de connexions plus ou moins fixes. Par exemple, il semble que la seule sorte de faits qui pourraient expliquer une certaine nappe de neurones comme « s'occupant » de la couleur soient des faits qui portent sur ses connexions propres, aussi indirectes soient-elles, aux cellules en cône de la rétine qui sont sensibles au plus haut point à différentes fréquences de lumière. À partir du moment où l'on a établi cette identité fonctionnelle, ces connexions pourraient être coupées (tout comme elles sont quelquefois rendues aveugles à l'âge adulte) sans que les spécialistes perdent (totalement) leur pouvoir de représenter (ou encore de « s'occuper ») de la couleur. Mais sans connexions causales de ce genre il est difficile de voir ce qui pourrait donner aux spécialistes un rôle dans la saisie d'un contenu spécifique⁴. Il semble donc que le cortex soit (largement) composé d'éléments dont les pouvoirs représentationnels plus ou moins fixes sont le résultat de leur localisation fonctionnelle dans l'ensemble du réseau. Ils correspondent à la manière dont les membres de la Chambre des Représentants représentent des districts : en transportant de l'information à partir de sources auxquelles ils sont liés spécifiquement (on peut découvrir que la plupart des transactions téléphoniques dans leurs bureaux à Washington se reliaient à leurs régions d'origine, par exemple). Or imaginez que les membres de la Chambre des Représentants soient assis sur un groupe de sièges dans un stade et qu'ils représentent le message important « La vitesse tue ! » en portant individuellement de grandes cartes de couleur au-dessus de leur tête, en épelant le message en lettres géantes visibles de l'autre côté du stade. Ce seraient des pixels vivants, donc, et leurs relations à leurs propres circonscriptions électorales ne joueraient pas de rôle dans leur contribution à la représentation du groupe. Certains modèles de recrutement cortical suggèrent fortement que *quelque chose comme* ce rôle représentatif secondaire doit être possible. Par exemple, il est tentant de supposer que les contenus informationnels portant sur un sujet particulier peuvent surgir dans une nappe de neurones spécialisés, et être ensuite, pour ainsi dire, propagés à travers les régions corticales, en exploitant la variabilité de ces régions sans

engager la sémantique spécialisée des unités qui y résident. Supposez, par exemple, qu'un changement soudain survienne dans l'angle visuel gauche supérieur du monde visuel d'une personne. Comme on peut s'y attendre, on peut repérer la naissance de l'excitation du cerveau dans les parties du cortex visuel qui représentent (à la façon de la Chambre des Représentants) divers traits des événements de l'angle visuel supérieur gauche de la vision, mais ces points chauds deviennent immédiatement des sources d'activation répandue, impliquant des agents corticaux qui ont d'autres circonscriptions. Si cette extension d'excitation sur des aires du cortex n'est pas simplement une fuite ou du bruit, si elle joue un rôle crucial quelconque dans l'élaboration et dans la facilitation de l'édition d'une version ou d'un fragment narratif, les agents ainsi recrutés doivent jouer un rôle tout à fait différent de celui qu'ils jouent quand ils sont à leur source d'ancrage ⁵.

Il n'est pas surprenant que nous n'ayons pas encore de bons modèles de ce genre de fonctionnalité multiple (les seules spéculations plausibles que j'aie lues sur ce sujet sont celles de Minsky, dans *La Société de l'esprit*). Comme nous l'avons noté au chapitre 6, les ingénieurs humains, dotés de capacités imparfaites de prévision, s'entraînent à concevoir des systèmes dans lesquels chaque élément joue un rôle unique, bien isolé des interférences venues de l'extérieur, afin de minimiser les conséquences dévastatrices d'effets secondaires imprévus. Dame Nature, quant à elle, ne s'inquiète pas de prévoir les effets secondaires ; elle peut ainsi capitaliser les effets secondaires ingénieux qui sont bénéfiques quand ils se montrent – un beau jour. Il est probable que le caractère opaque de la décomposition fonctionnelle dans le cortex a jusqu'à présent résisté aux efforts des neuro-anatomistes provient parce qu'ils se trouvent eux-mêmes constitutionnellement incapables de faire des hypothèses qui assignent des rôles multiples aux éléments disponibles. Il est des romantiques – que le philosophe Owen Flanagan (1991) appelle les Nouveaux Adorateurs de Mystère – qui ont avancé l'idée que le cerveau rencontre un obstacle insurmontable à comprendre sa propre organisation (Nagel, 1986, et Mc Ginn, 1990). Je ne forme aucune hypothèse de ce genre, mais je suppose seulement qu'il se révèle terriblement difficile – mais pas impossible – de se faire une idée de la façon dont le cerveau fonctionne, en partie parce qu'il a été construit par un processus qui peut s'appuyer sur une fonc-

tionnalité multiple, superposée, qu'il est systématiquement difficile de discerner du point de vue de l'ingénierie inversée.

C'est le genre de problèmes qui provoquent des allusions vagues, quand ils sont ne serait-ce que pris en compte. Certains sont tentés de rejeter d'emblée l'idée d'une dualité spécialiste/généraliste de ce genre – non parce qu'ils peuvent démontrer qu'elle est erronée, mais parce qu'ils ne parviennent pas à imaginer comment la modéliser, ce qui leur fait espérer à très juste titre qu'ils n'auront jamais à le faire. Mais, une fois que cette perspective se trouve ouverte, elle donne au moins quelques nouveaux indices aux chercheurs. Les neurophysiologistes ont identifié (ou tenté d'identifier) des mécanismes dans les neurones tels que les récepteurs NMDA et les synapses de von der Malsburg (1985) qui sont des candidats appropriés pour jouer un rôle comme modulateurs rapides de connectivité entre les cellules. De telles portes pourraient permettre la formation rapide d'« assemblées » transitoires, qui pourraient être superposées sur les réseaux sans requérir les altérations des forces synaptiques à long terme qui sont en général supposées être la glu qui fait tenir ensemble les assemblées permanentes de la mémoire à long terme. (Pour des spéculations nouvelles en ce sens, voir Flohr, 1990.)

À une échelle plus vaste, les neuro-anatomistes ont rempli la carte des connexions dans le cerveau. Ils ont établi ainsi non seulement quelles aires étaient actives dans quelles circonstances, mais aussi commencé à montrer quelles sortes de contributions elles apportaient. On a supposé que plusieurs aires jouaient un rôle crucial dans la conscience. On sait depuis longtemps que la *formation réticulée* dans le cerveau moyen et le thalamus situé au dessus jouent un rôle crucial dans l'éveil du cerveau – à partir du sommeil, par exemple, ou en réponse à des événements nouveaux ou à des dangers ; à présent que l'on a dressé la carte des chemins, on peut formuler des hypothèses plus détaillées et les tester. Crick (1984), par exemple, propose que les branches qui rayonnent à partir du thalamus vers toutes les parties du cortex l'adaptent bien à son rôle de « projecteur », éveillant ou excitant différenciellement diverses aires spécialisées, et les recrutant pour divers objectifs ⁶. Baars (1988) a élaboré une idée similaire : le ERTAS ou Système d'Activation Thalamique Réticulaire Étendu (*Extended Reticular Thalamic Activating System*). Il serait assez facile d'incorporer ce genre d'hypothèse dans notre analyse anatomiquement neutre de la compétition entre les coalitions de

spécialistes, à condition que nous ne tombions pas dans le piège qui consiste à imaginer qu'il existe un Chef thalamique *compre-nant* les événements gérés par les diverses parties du cerveau avec lesquels il est « en communication ».

De même, on sait que les *lobes frontaux* du cortex, partie du cerveau qui est le plus nettement agrandie chez l'*Homo sapiens*, sont impliqués dans le contrôle à long terme, et dans l'organisation et la sériation du comportement. Des lésions dans les diverses régions des lobes frontaux produisent typiquement des symptômes conflictuels tels que la distractibilité par opposition à l'inaptitude trop attentive à se désexciter et l'impulsivité ou l'inaptitude à suivre des schèmes d'action qui requièrent qu'on retarde leur gratification. Il est donc tentant d'installer le Chef dans les lobes frontaux, et plusieurs modèles vont dans cette direction. L'un des plus sophistiqués est celui de Norman et Shallice (1985) : le Système de Supervision Attentionnelle, qu'ils localisent dans le cortex préfrontal et auquel ils donnent la responsabilité particulière de résoudre les conflits quand des bureaucraties subsidiaires ne parviennent pas à coopérer. Une fois encore, c'est une chose que de trouver une localisation pour des processus qui sont cruciaux dans le contrôle de ce qui va se passer ensuite, et c'en est une autre que de localiser le Chef ; quiconque part à la chasse de l'écran frontal sur lequel le Chef garde la trace des projets qu'il contrôle part à la chasse au Dahu (Fuster, 1981 ; Calvin, 1989a.)

Mais si nous écartons ces images tentantes, nous avons à chercher d'autres façons de penser les contributions de ces aires, et on est encore à court d'idées, en dépit des progrès récents. Ce n'est pas que nous n'ayons pas idée de la nature de la machinerie ; le problème tient bien plus au fait que nous ne disposons pas d'un modèle computationnel de ce que fait la machinerie, et comment. Nous en restons donc toujours aux métaphores et aux allusions. Mais il ne faut pas mépriser cette étape ; il faut en passer par là avant de parvenir à des modèles plus explicites.

Les pouvoirs de la machine joycienne

Selon l'esquisse de théorie que j'ai présentée, il y a compétition entre de nombreux événements dotés de contenu qui se pro-

duisent dans le cerveau et c'est un sous-ensemble de ces événements, produit d'une sélection, qui « gagne ». Cela veut dire que ces événements parviennent à répercuter des effets continus de diverses sortes. Certains, s'unissant avec des démons du langage, contribuent à des énonciations ultérieures, à la fois des énoncés adressés tout haut à d'autres et des énoncés silencieux adressés à soi-même. Certains prêtent leur contenu à d'autres formes d'autostimulation ultérieures, comme se faire des dessins à soi-même. Le reste des événements meurent presque immédiatement, ne laissant que de faibles traces – des preuves momentanées – de leur passage. À quoi cela sert-il, pourriez-vous vouloir demander, que certains contenus gagnent leur entrée dans le cercle enchanté – et qu'y a-t-il de si enchanté dans ce cercle ? La conscience est supposée être quelque chose de terriblement spécial. Qu'y a-t-il de si spécial dans le fait de se trouver avancé jusqu'au prochain round d'un tel cycle d'autostimulation ? En quoi cela aide-t-il ? Est-ce que des pouvoirs presque magiques se trouvent conférés aux événements qui se produisent dans de tels mécanismes ?

J'ai évité de soutenir qu'une forme particulière de victoire dans ce tourbillon de luttes revient à les promouvoir à la conscience. En réalité, j'ai insisté sur le fait qu'il n'est pas possible de justifier l'existence d'une frontière entre des événements qui appartiendraient de façon définie à la conscience et des événements qui se tiendraient à jamais « en dehors » ou « en deçà » de la conscience. (Voir Allport, 1988, qui donne d'autres arguments en faveur de cette position.) Néanmoins, si ma théorie de la machine joycienne doit vraiment éclairer la nature de la conscience, il faut bien qu'il y ait quelque chose de remarquable quant à certaines au moins des activités de cette machine, parce qu'on ne peut nier que la conscience soit, intuitivement parlant, quelque chose de spécial.

Il est difficile de s'attaquer à ces questions familières sans tomber dans le piège suivant : penser que nous devons avoir une idée de ce à quoi sert la conscience, en sorte que nous puissions ensuite demander si les mécanismes proposés parviendraient à accomplir *cette* fonction – quelle que soit la manière dont nous en déterminions la nature.

Dans son livre influent intitulé *Vision*, le chercheur en neurosciences et en IA David Marr a proposé qu'on distingue trois niveaux d'analyse dans toute tentative d'explication des phéno-

mènes mentaux. Le niveau « supérieur » le plus abstrait, le niveau computationnel, est une analyse du « *problème* (les italiques sont de moi) comme tâche de traitement de l'information », alors que le niveau du milieu, le niveau algorithmique, recouvre l'analyse des processus effectifs par lesquels cette tâche de traitement de l'information est accomplie. Le niveau le plus bas, le niveau physique, fournit une analyse de la machinerie neuronale et montre comment elle exécute les algorithmes décrits au niveau du milieu, accomplissant ainsi la tâche décrite abstraitement au niveau computationnel ⁷.

On peut aussi utiliser les trois niveaux de Marr pour décrire des choses qui sont bien plus simples que des esprits, et on peut se faire une idée de la différence entre les niveaux en regardant comment ils s'appliquent à quelque chose de simple, comme un boulier. Sa tâche computationnelle est de faire de l'arithmétique : il s'agit de fournir une sortie correcte pour n'importe quel problème arithmétique qui lui est fourni comme entrée à ce niveau ; par conséquent, un boulier et une calculatrice de poche sont identiques : ils sont conçus pour accomplir la même « tâche de traitement d'information ». La description arithmétique du boulier est ce que vous apprenez quand vous apprenez à le manipuler – la recette pour mouvoir les boules quand vous additionnez, soustrayez, multipliez et divisez. Sa description physique dépend de ce dont il est fait : ce peut être des boules en bois enfilées sur un fil à l'intérieur d'un cadre, ou bien ce peut être des jetons de poker alignés le long des rainures du parquet, ou quelque chose que l'on fait avec un crayon et une bonne gomme sur un morceau de papier quadrillé.

Marr recommandait une modélisation des phénomènes psychologiques aux trois niveaux d'analyse et il insistait particulièrement sur l'importance d'une description claire du niveau supérieur computationnel avant de se précipiter tête baissée vers la modélisation des niveaux inférieurs ⁸. Son œuvre propre sur la vision manifestait brillamment le pouvoir de cette stratégie, et d'autres chercheurs en ont depuis fait bon usage dans d'autres domaines. Il est tentant d'appliquer les mêmes trois niveaux d'analyse à la conscience dans son ensemble, et certains ont succombé à la tentation. Comme nous l'avons vu au chapitre 6, cependant, on risque de simplifier à l'excès. En posant la question de savoir quelle est *la* fonction de la conscience ?, on présuppose qu'il existe une unique « tâche de traitement de l'information »

(aussi complexe soit-elle) que la machinerie neuronale de la conscience est bien construite pour accomplir – par l'évolution, pense-t-on. Cela peut nous conduire à négliger des possibilités importantes : le fait que certains traits de la conscience ont des fonctions multiples ; que certaines fonctions ne sont que mal desservies par les traits existants, en raison des limitations historiques de leur développement ; que certains traits n'ont pas de fonction du tout – du moins pas de fonction qui puisse *nous* procurer un avantage. En faisant attention, par conséquent, d'éviter ces simplifications, passons en revue les *pouvoirs* (pas nécessairement les fonctions) des mécanismes décrits dans mon croquis schématique.

En premier lieu, comme nous l'avons vu au chapitre 6, il existe des problèmes significatifs d'autocontrôle dus à la prolifération systématique de spécialistes qui entrent conjointement en activité ; l'une des tâches fondamentales que réalise la machine joycienne consiste à régler les disputes, à adoucir les transitions entre régimes et à empêcher les *coups d'État* inopinés en mettant en place les « bonnes forces ». Des tâches simples ou qui ont fait l'objet d'un apprentissage répété sans compétition sérieuse peuvent être exécutées de façon routinière sans que soient enrôlées d'autres forces, et par conséquent inconsciemment ; mais quand une tâche est difficile ou déplaisante elle requiert de la « concentration », quelque chose que « nous » accomplissons sans l'aide d'admonestations que nous nous ferions à nous-mêmes ou de divers trucs mnémoniques, de répétition (Margolis, 1989) et d'autres automanipulations (Norman and Shallice, 1985). Il arrive souvent que nous trouvions utile de penser tout haut : c'est un retour aux stratégies frustes mais efficaces dont nos pensées privées sont de jolis descendants.

De telles stratégies d'autocontrôle nous permettent de gouverner nos propres processus perceptuels de manière à obtenir de nouvelles opportunités. Comme le psychologue Jeremy Wolfe (1990) le fait remarquer, bien que nos systèmes visuels soient construits de façon innée pour détecter certaines sortes de choses – celles qui « se pointent » quand nous « jetons un œil » – il existe d'autres sortes de choses que nous ne pouvons identifier que si nous pouvons les *chercher*, délibérément, dans le cadre d'une politique reposant sur un acte d'autoreprésentation. Une tache rouge au sein d'une grande quantité de taches vertes nous frappera comme le ferait un pouce douloureux (en fait, elle nous

frappera comme le ferait une framboise mûre au milieu des feuilles), mais si vos projets sont tels que vous cherchez une tache rouge dans une foule de taches de multiples couleurs, il vous faudra *vous fixer à vous-même* une tâche de recherche en série. Et si vous avez le projet de trouver un confetti carré rouge au milieu de ses voisins de toutes les couleurs et de toutes les formes (ou de répondre à la question « Où est Waldo ? » (Handford, 1987) dans les petits dessins-énigmes populaires), la tâche de recherche en série deviendra un projet particulièrement développé, méthodique, qui supposera beaucoup de contrôle de soi.

Ces techniques de représentation (de choses à nous-mêmes) nous permettent d'être des gouvernants ou des directeurs d'une façon dont aucune autre créature n'est capable. Nous pouvons mettre en place nos politiques bien à l'avance, grâce à notre capacité à former des hypothèses et à tramer des scénarios. Nous pouvons renforcer nos résolutions de nous engager dans des projets désagréables ou à long terme grâce à des habitudes de remémoration, et en répétant les bénéfices et les coûts attendus des politiques que nous avons adoptées. Chose plus importante encore, cette pratique de répétition crée un souvenir des voies par lesquelles nous sommes parvenus à l'endroit où nous nous trouvons (ce que les psychologues appellent *mémoire épisodique*), en sorte que nous puissions nous expliquer à nous-mêmes, quand nous nous retrouvons coincés, quelles sortes d'erreurs nous avons faites (Perlis, 1991). Au chapitre 6, nous avons vu comment le développement de ces stratégies avait permis à nos ancêtres de continuer leur recherche dans le futur et qu'ils avaient acquis ce pouvoir supplémentaire d'anticipation en partie en raison d'un plus grand pouvoir de réminiscence : ils étaient capables de revenir plus loin en arrière à leurs opérations récentes pour voir où ils avaient fait des erreurs. « Je ne dois pas refaire *cela* » est le refrain de toute créature qui apprend à partir de l'expérience, mais nous pouvons apprendre à envoyer nos *cela* plus loin et plus lucidement en arrière que n'importe quelle créature, grâce à notre habitude de conserver des données en mémoire, ou, plus exactement, grâce à nos habitudes d'autostimulation qui ont, parmi leurs nombreux effets, permis l'augmentation de la réminiscence.

Mais cette augmentation de la charge de la mémoire n'est que l'un des effets utiles de ces habitudes. L'effet de diffusion est tout aussi important (Baars, 1988) : il crée une sorte de forum ouvert,

permettant à *n'importe laquelle* des choses qui ont été apprises d'apporter une contribution à *n'importe quel* problème présent. Baars soutient que cette accessibilité mutuelle des contenus fournit le *contexte* sans lequel le sujet ne donnerait pas (ou ne pourrait pas donner) de sens à des événements survenant « dans la conscience ». Les contenus qui composent le contexte environnant ne sont pas eux-mêmes toujours conscients – en fait, ils ne sont en général pas accessibles du tout, en dépit du fait qu'ils soient activés – mais les connexions entre eux et les contenus qui peuvent se révéler dans des comptes rendus verbaux sont ce qui assure ce que nous pourrions appeler leur signification « consciemment appréhendée ».

Ray Jackendoff (1987) soutient dans le même esprit que les niveaux les plus élevés d'analyse auxquels puisse atteindre le cerveau – par quoi il entend les plus abstraits – ne sont *pas* accessibles à l'expérience, en dépit du fait qu'ils rendent l'expérience possible, en lui donnant un sens. Son analyse nous donne donc un antidote utile contre une autre incarnation du Théâtre Cartésien comme « sommet » ou « pointe de l'iceberg » (voici un bon exemple, emprunté au neuropsychologue Roger Sperry : « Dans une position de commandement en chef aux niveaux les plus élevés de la hiérarchie de l'organisation cérébrale, les propriétés subjectives[...] exercent un contrôle sur les activités biophysiques et chimiques situées aux niveaux subordonnés » (1977, p. 171)).

Un très petit nombre de philosophes, en particulier ceux qui sont influencés par l'école de la phénoménologie husserlienne (Dreyfus, 1979 ; Searle, 1983), ont souligné l'importance de cet « arrière-plan » d'expérience consciente, mais ils l'ont en général décrit comme un trait mystérieux et récalcitrant, qui défie toute explication mécanique, plutôt que comme la clef, ainsi que le suggèrent Baars et Jackendoff, d'une théorie computationnelle de ce qui se produit dans la conscience. Ces philosophes ont supposé que la conscience est la *source* d'une espèce particulière d'« intentionnalité intrinsèque », mais comme le philosophe Robert Van Gulick l'a noté, cela ne fait que nous ramener en arrière.

« L'expérience de compréhension au niveau personnel [...] n'est pas une illusion. Moi, le sujet personnel de l'expérience, comprends, je peux faire toutes les connexions nécessaires avec l'expérience, appelant des représentations à se relier immédiatement les unes aux autres. Le fait que ma capacité soit le résultat du fait que je

sois composé d'un système organisé de sous-composants qui produisent mon flux ordonné de pensées ne gêne pas ma capacité. Ce qui est illusoire ou erroné n'est que la thèse selon laquelle je suis un moi substantiel distinct qui produit ces connexions en vertu d'une forme totalement non comportementale de compréhension » (Van Gulick, 1988, p. 96).

N'importe laquelle des choses que vous avez apprises peut contribuer à n'importe laquelle des choses auxquelles vous êtes présentement confronté. Du moins est-ce la situation idéale. Fodor (1983) appelle ce trait *isotropie* : c'est le pouvoir, comme l'aurait dit Platon, de faire venir les bons oiseaux, ou tout au moins de les siffler chaque fois qu'on en a besoin. Cela a l'air magique, mais comme le sait tout magicien de théâtre, l'apparence de magie est renforcée par le fait qu'on peut s'attendre à ce qu'un auditoire exagère le besoin d'expliquer tel ou tel phénomène. Nous pouvons de prime abord avoir l'air d'être idéalement isotropiques, mais nous ne le sommes pas. Quand nous réfléchissons à froid, nous nous rappelons toutes les occasions où nous avons été lents à reconnaître le sens de nouvelles données. Pensez à l'exagération comique de ce trait : les actions « en deux temps » (Neisser, 1988). Il nous arrive même de scier la branche sur laquelle nous sommes assis ou de craquer une allumette pour y voir mieux dans un réservoir à essence ?

Les magiciens de théâtre savent qu'une collection de trucs faciles suffiront souvent à produire de la « magie ». Dame Nature aussi, elle qui est la plus grande pourvoyeuse de gadgets. La recherche en intelligence artificielle a exploré l'espace des trucs possibles, en cherchant « un tas[...] d'heuristiques bien coordonnées et rapidement déployées » (Fodor, p. 116) qui pourraient fournir le degré d'isotropie que nous, êtres humains, manifestons. Des modèles comme ACT* et Soar – et beaucoup d'autres visions envisagées en IA – sont prometteurs mais incomplets. Certains philosophes, en particulier Dreyfus, Searle, Fodor et Putnam (1988), sont sûrs que cette idée de l'Esprit comme Gadget est fausse ; ils ont essayé de construire des arguments destinés à prouver l'impossibilité de la tâche (Dennett, 1988b, 1991c). Fodor, par exemple, fait remarquer que bien que les systèmes à objectifs limités puissent être innés, dans un système à objectifs généraux qui puisse répondre avec versatilité à n'importe quelle nouvelle donnée qui survienne, « ce peut être la connectivité *instable, instantanée, qui compte* » (p. 118). Il désespère qu'on puisse jamais

fournir une théorie de cette connectivité, mais il n'est pas seulement pessimiste : c'est un désespoir *de principe* (un joli truc). Il a raison de dire que nous devrions nous attendre à ce que notre approximation de l'isotropie soit due à notre logiciel, et pas à notre câblage inné, mais son argument contre l'hypothèse du « sac à malices » suppose que nous sommes meilleurs pour « considérer toutes choses » que nous ne le sommes en fait. Nous sommes bons, mais nous ne sommes pas fantastiques. Les habitudes d'autostimulation que nous développons font de nous des individus capables d'exploiter avec astuce nos ressources durement gagnées ; nous ne parvenons pas toujours à faire chanter le bon oiseau au bon moment, mais nous y arrivons assez souvent pour que nous puissions nous retrouver en assez bonne compagnie.

Est-ce bien une théorie de la conscience ?

Jusqu'à présent, j'ai été timide au sujet de la conscience. J'ai pris soin d'éviter de vous dire ce que ma théorie dit sur la *nature* de la conscience. Je n'ai pas déclaré que tout ce qui exemplifie une machine joycienne est conscient et que n'importe quel état particulier d'une telle machine virtuelle était un état conscient. La raison de ma réticence était tactique : je voulais éviter de polémiquer sur la nature de la conscience avant d'avoir eu la possibilité de montrer qu'au moins une grande quantité des pouvoirs présumés de la conscience pouvaient s'expliquer par les pouvoirs de la machine joycienne, *que la machine joycienne dote ou non le matériel qui l'héberge d'états conscients*.

Ne pourrait-il exister un être inconscient avec un espace interne global de travail dans lequel des démons diffuseraient des messages à d'autres démons, formant des coalitions et tout le reste ? Dans ce cas, le pouvoir étonnant qu'ont les êtres humains d'ajuster rapidement et avec versatilité leurs états mentaux en réponse à presque n'importe quelle circonstance, aussi nouvelle soit-elle, ne devrait rien à la conscience elle-même ; il serait dû simplement à l'architecture computationnelle qui rend possible cette intercommunication. Si la conscience est quelque chose qui se tient au-dessus et au-delà de la machine joycienne, je n'ai toujours pas donné une théorie de la conscience, même si j'ai répondu à d'autres questions délicates.

Tant que mon ébauche de théorie n'était pas complète, je devais différer une réponse à ce doute, mais il faut maintenant y aller carrément et affronter la conscience elle-même, l'ensemble du merveilleux mystère. Je déclare donc ici que OUI, ma théorie est une théorie de la conscience. Quiconque ou quoi que ce soit qui dispose d'une machine virtuelle de ce type comme système de contrôle est conscient au sens le plus plein du terme ; il est conscient *parce qu'il a une telle machine virtuelle* ¹⁰.

Je suis prêt à présent à répondre aux objections. Nous pouvons commencer par nous atteler à la question qui a été laissée sans réponse deux paragraphes plus haut. Quelque chose d'inconscient – un zombie par exemple – ne pourrait-il disposer d'une machine joycienne ? Cette question implique une objection qui surgit si souvent à des moments comme celui-ci que le philosophe Peter Bieri l'a appelée *le moulin à prières tibétain*. Elle ne cesse de se poser, encore et toujours, quelle que soit la théorie que l'on ait avancée :

Tout ça c'est bien joli, tous ces détails fonctionnels sur la façon dont le cerveau fait ceci ou cela, mais je peux imaginer que *tout cela* se produit dans une entité sans qu'il y ait la moindre occurrence d'une véritable conscience.

Une bonne réponse à cette question, mais qu'on entend rarement, est : Oh ? vraiment ? Comment le savez-vous ? Comment savez-vous que vous avez imaginé « tout cela » avec assez de détails, et en étant assez attentif à toutes les implications ? Qu'est-ce qui vous fait penser que votre thèse est une prémisse conduisant à une conclusion intéressante ? Voyons combien nous serions peu impressionnés si un *vitaliste* contemporain nous disait :

Tout ça c'est bien joli, tous ces machins sur l'ADN, les protéines et ce genre de choses, mais je peux imaginer que l'on découvre une entité qui aurait la même apparence et agirait de la même façon qu'un chat, jusqu'au sang dans ses veines et l'ADN de ses « cellules », mais qui ne serait pas « vivant ». (Est-ce que je peux vraiment l'imaginer ? Sûrement : le voilà, miaulant, et voici que Dieu me murmure dans l'oreille : « Ce n'est pas vivant ! C'est juste un machin-chose-ADN mécanique ! » Dans mon imagination je Le crois.)

Je suis sûr que personne ne pense que c'est un bon argument en faveur du vitalisme. Cet effort d'imagination ne compte pas. Pourquoi pas ? Parce qu'il est trop faiblard pour peser lourd

contre l'analyse de la vie que nous donne la biologie contemporaine. La seule chose que montre cet « argument » est que vous pouvez ignorer « tout cela » et vous raccrocher à votre conviction si vous êtes déterminé à le faire. Est-ce que le moulin à prières tibétain est un argument meilleur contre la théorie que j'ai esquissée ?

Nous sommes à présent en mesure, grâce aux pouvoirs d'imagination déployés dans les chapitres précédents, de déplacer le poids de la preuve. Le moulin à prières tibétain (et il y a plusieurs variantes importantes, comme nous le verrons) descend en droite ligne du fameux argument de Descartes (voir chapitre 1), dans lequel celui-ci prétend être capable de concevoir *clairement et distinctement* que son esprit est distinct de son corps. La force de ce genre d'argument dépend de façon cruciale de la nature des critères de conception que l'on a et de leur degré d'élévation. Certaines personnes peuvent soutenir qu'elles peuvent clairement et distinctement concevoir un nombre premier qui soit le plus grand ou un triangle qui n'est pas une figure rigide. Ils ont tort – ou du moins, ce qu'ils font quand ils disent qu'ils conçoivent ces choses ne devrait pas être considéré comme un signe de ce qui est possible. Nous sommes à présent en position d'imaginer « tout ça » de façon plus détaillée. Pouvez-vous *réellement* imaginer un zombie ? Le seul sens dans lequel c'est « évident » que vous puissiez le faire ne peut menacer ma théorie, et si l'on adopte un sens plus fort, moins évident, il faut démontrer qu'il est possible.

Les philosophes n'ont pas, en règle générale, exigé de telles choses. Les expériences de pensée les plus influentes de la philosophie de l'esprit récente ont toutes impliqué qu'on invite l'auditoire à imaginer un état de choses particulièrement alambiqué ou stipulé, puis – sans vérifier à proprement parler que ce tour d'imagination a été effectivement accompli – on a invité l'auditoire à « remarquer » diverses conséquences de la fiction ainsi inventée. Ces « pompes à intuition », comme je les appelle, sont souvent des dispositifs diaboliquement habiles. Elles méritent leur célébrité ne serait-ce qu'en raison de la séduction qu'elles exercent.

Dans la Troisième Partie, nous les reprendrons toutes, en développant au fur et à mesure notre théorie de la conscience. À partir de notre nouvelle perspective, nous pouvons nous demander où est le tour de passe-passe qui trompe l'auditoire – et les

illusionnistes. En cours de route nous aiguïserons nos pouvoirs d'imagination. Parmi les arguments fameux que nous rencontrerons ne figurent pas seulement la Possibilité Présumée des Zombies, mais le Spectre Inversé, Ce que Marie la Neurophysiologiste des Couleurs ne Sait Pas sur les Couleurs, la Chambre Chinoise, et Quel Effet Cela Fait d'Être une Chauve-Souris.

NOTES

1. En français dans le texte (NdT).

2. Edelman (1989) est un théoricien qui a essayé de mettre tout ensemble, des détails de la neuro-anatomie à la psychologie cognitive et aux modèles computationnels, ainsi que jusqu'aux controverses philosophiques les plus absconses. Le résultat est un échec instructif. Il montre en grand détail comment il faut répondre à de nombreuses sortes de questions différentes avant de pouvoir soutenir qu'on a fourni une théorie complète de la conscience, mais il montre aussi qu'aucun théoricien ne peut apprécier les subtilités des problèmes auxquels on s'adresse dans différents domaines. Edelman a mésinterprété, puis rejeté abruptement, les travaux de nombre de ses alliés potentiels. Il a ainsi isolé sa théorie de la sorte d'attention bienveillante et informée qu'elle requiert si on doit la sauver de ses erreurs et de ses défauts. Cela soulève la question parallèle de savoir si moi aussi j'ai sous-estimé certains des travaux avec lesquels je suis en désaccord dans ces pages ; sans aucun doute je l'ai fait, et j'espère que ceux dont j'ai mal représenté les rejetons de leurs cerveaux essaieront (une fois encore) d'expliquer ce que j'ai manqué.

3. Les fonctionnalistes ont pris l'habitude de faire de la « boîtologie » – dessiner des diagrammes qui installent les diverses fonctions dans des boîtes séparées, tout en niant explicitement que ces boîtes aient la moindre signification anatomique. (Je suis moi-même coupable d'avoir initié, et encouragé, cette pratique ; voir les figures dans *Brainstorms*, chapitres 7, 9, et 11.) Je pense toujours qu'« en principe c'est une bonne tactique », mais en pratique elle tend à rendre le fonctionnaliste aveugle aux différentes décompositions possibles de fonctions, et en particulier à la possibilité qu'il existe de nombreuses fonctions superposées les unes sur les autres. L'image de la séparation mentale entre la mémoire de travail et la mémoire à long terme – image aussi ancienne que la volière de Platon – joue un rôle non trivial dans la façon dont les théoriciens conçoivent les tâches de la cognition. Un exemple frappant : « Le besoin de symboles intervient parce qu'il n'est pas possible que toute la structure impliquée dans un calcul soit assemblée par avance dans le site physique de la computation. Ainsi il est nécessaire de voyager vers d'autres parties (distales) de la mémoire pour obtenir la structure additionnelle » (Newell, Rosenbloom and Laird, 1989, p. 105). Cela conduit très directement à l'image de *symboles amovibles*, puis, chez ceux qui s'enthousiasment pour cette image sans recul critique, au scepticisme quant à toutes les architectures connexionnistes, sur la base du fait que les éléments de ces architectures qui s'approchent le plus des symboles – les nœuds qui d'une manière ou d'une autre ancrent la sémantique du système – sont inamovibles dans leur réseau d'interconnexions. Voir, par exemple, Fodor et Pylyshyn (1988). Le problème des éléments sémantiques fixes par opposition à ceux qui sont amovibles n'est que l'une des façons de considérer l'un des problèmes fondamentaux non résolus des sciences cognitives. Ce n'est probablement pas comme cela qu'il faut le considérer, mais il ne disparaîtra pas tant qu'il ne sera pas remplacé par une conception meilleure, ancrée dans un accord positif – par

opposition à un rejet hystérique –, des faits fondamentaux de la neuro-anatomie fonctionnelle.

4. En d'autres termes, les attrait d'une « théorie causale de la référence » sont aussi évidents pour les chercheurs en sciences cognitives qu'ils le sont pour les philosophes.

5. Fodor note une variante de ce problème dans sa discussion de « avoir un concept » (1990, p. 80-81).

6. Les théories de l'attention reposant sur l'image du projecteur sont connues depuis des années. Les théories simples font l'erreur de supposer trop littéralement que ce que le projecteur illumine différenciellement ou excite à un moment donné est une région de l'*espace visuel* – exactement de la manière dont un spot dans un théâtre peut illuminer une région de la scène à un moment donné. Les théories du projecteur qui sont plus défendables (mais aussi plus impressionnistes) insistent sur le fait que ce qui est excité différenciellement est une portion de l'*espace conceptuel* ou *sémantique* (imaginez, si vous le pouvez, un spot de théâtre qui puisse éclairer seulement les *Capulets*, ou seulement l'ensemble des *amoureux*). Voir Allport (1989) sur les difficultés des théories fondées sur l'image du projecteur.

7. Des distinctions étroitement liées à celle-ci sont ma trinité de la posture intentionnelle, de la posture du plan et de la posture physique (Dennett, 1971), et l'identification par Allan Newell (1982) du « niveau de la connaissance » situé au-dessus du « niveau du système des symboles physiques ». Voir Dennett, 1987a, 1988e, et Newell, 1988.

8. Comme Marr le notait : « Il devient possible, en séparant les explications en différents niveaux, de faire des énoncés explicites quant à ce qui est calculé et pourquoi et de construire des théories énonçant que ce qui est calculé est optimal en un sens quelconque ou est garanti fonctionner correctement » (p. 19). Pour plus de détails sur les avantages et les pièges de cette ingénierie à l'envers voir Dennett (1971, 1983, 1987a, 1988d), Ramachandran (1985).

9. Dans *Minimal Rationality* (1986), le philosophe Christopher Cherniak analyse les perspectives et les limites des processus déductifs rendus possibles par ce forum ouvert. Voir aussi Stalnaker (1984).

10. Jackendoff (1987) adopte une tactique quelque peu différente. Il divise le problème esprit/corps en deux, et destine sa théorie à répondre à la question de savoir comment l'esprit *computationnel* s'adapte au corps ; cela le laisse avec un « problème corps/esprit » non résolu : quelle relation y a-t-il entre l'esprit phénoménologique et l'esprit computationnel ? Plutôt que de concéder que ce second problème est un mystère résiduel, je propose de montrer que le modèle des Versions Multiples, joint à la méthode de l'hétérophénoménologie, dissout les deux problèmes en même temps.

TROISIÈME PARTIE

**Les problèmes philosophiques
de la conscience**

Chapitre 9

Montrer et dire

La rotation d'images dans l'œil de l'esprit

Le premier défi à relever, avant que nous en venions aux expériences de pensée, provient de certaines expériences réelles qui pourraient sembler réhabiliter le Théâtre Cartésien. Certaines des recherches parmi les plus excitantes et les plus ingénieuses des sciences cognitives ces vingt dernières années ont porté sur l'aptitude humaine à manipuler des images mentales. Elles ont pour origine le travail classique du psychologue Roger Shepard (Shepard and Meltzer, 1971) sur la *vitesse de rotation mentale* de figures comme celles-ci.

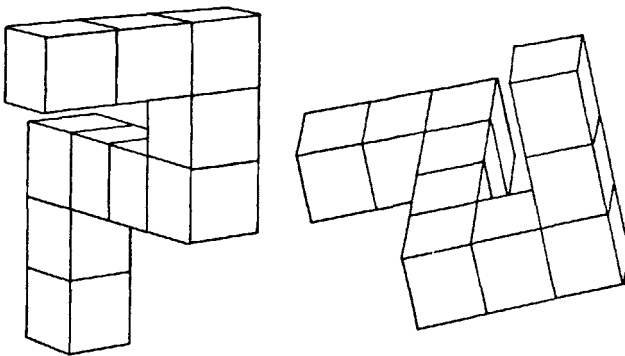


Figure 9.1

Dans les expériences originales, on montrait aux sujets des paires de dessins de ce genre, et on leur demandait si ces paires étaient ou non des points de vue différents de la même forme. Dans ce cas, comme vous pouvez le déterminer rapidement, la réponse est oui. Une réponse typique est : « J'ai fait tourner l'une des images dans l'œil de mon esprit, et je l'ai superposée sur l'autre. » Shepard variait les distances de rotation angulaire entre les paires de figures – certaines paires n'étaient qu'à quelques degrés de l'alignement – et mesurait le temps de réponse moyen des sujets aux différents dispositifs. À supposer qu'il existe quelque chose comme un processus de rotation effective de l'image dans le cerveau, cela devrait prendre en gros deux fois plus de temps pour faire tourner une image mentale sur 90 degrés que sur 45 degrés (si nous ignorons l'accélération et la décélération, en gardant la vitesse de rotation constante ¹). Les données de Shepard ont remarquablement bien confirmé cette hypothèse, dans nombre de cas. Des centaines d'expériences ultérieures, menées par Shepard et par d'autres, ont exploré le comportement de la machinerie de manipulation d'images du cerveau en détail et – pour formuler le consensus encore controversé avec autant de précautions que possible – il semble qu'il y ait ce que le psychologue Stephen Kosslyn (1980) appelle un « butoir (*buffer*) visuel » dans le cerveau et qu'il accomplisse des transformations par des processus qui sont fortement « imagés » – ou, pour utiliser le terme de Kosslyn, *quasi pictorial*.

Qu'est-ce que cela veut dire ? Les psychologues cognitifs ont-ils découvert que le Théâtre Cartésien existe finalement ? Selon Kosslyn, ces expériences montrent que les images sont assemblées en vue d'une présentation interne d'une manière très semblable à celle dont les images sur un CRT (un tube à rayons cathodiques comme un écran de télévision ou le moniteur d'un ordinateur) peuvent être créées à partir de dossiers dans la mémoire d'un ordinateur. À partir du moment où elles sont sur l'écran interne, les sujets peuvent les faire tourner, les rechercher, et les manipuler de diverses façons pour différentes utilisations. Kosslyn insiste, cependant, sur le fait que son modèle du CRT est une métaphore. Cela devrait nous rappeler les talents de « manipulation d'images » métaphorique dont disposait Shakey. Ce dernier n'avait certainement pas de Théâtre Cartésien dans son cerveau ordinateur. Pour avoir une image plus claire de ce qui doit en fait se produire dans un cerveau humain, nous pouvons partir

d'un modèle *non* métaphorique, beaucoup trop fort pour être correct, puis « soustraire » du modèle les propriétés indésirables les unes après les autres. En d'autres termes, nous prendrons la métaphore du CRT de Kosslyn, puis nous y introduirons petit à petit des limitations.

En premier lieu, considérons un système qui manipule réellement des images réelles, telles que les systèmes de graphisme sur ordinateur capables de créer des centaines de formes qui prolifèrent aujourd'hui : l'animation par ordinateur pour la télévision et les films, les systèmes qui rendent des objets-tridimensionnels en vues perspectives pour les architectes et les décorateurs d'intérieur, les jeux vidéo, et bien d'autres. Les ingénieurs appellent leurs versions des systèmes CAD, pour « conceptions aidées par ordinateur » (*Computer Aided Designs*). Les systèmes CAD révolutionnent l'ingénierie non seulement parce qu'ils rendent les avant-projets bien plus aisés, de la même façon que les traitements de texte rendent la saisie bien plus facile, mais parce que les ingénieurs peuvent aisément résoudre des problèmes et répondre à des questions qui seraient autrement très difficiles. Confronté au problème de Shepard (figure 9.1), un ingénieur pourrait répondre à la question avec l'aide d'un CAD en plaçant les images sur l'écran de son CRT et en faisant tourner littéralement l'une des images, puis en essayant de la surimposer sur les autres. Certains détails de ce processus sont importants.

Chaque objet représenté serait introduit dans la mémoire de l'ordinateur en tant qu'objet tridimensionnel *virtuel*, en le décomposant en une description de ses parties planes et de ses bords définis par des coordonnées *xyz*, chaque point occupé de l'espace virtuel étant un « triplet ordonné » de nombres stockés dans la mémoire de l'ordinateur. Le point de vue de l'observateur impliqué serait aussi introduit comme point dans le même espace virtuel, défini par son propre triplet ordonné *xyz*. Voici le diagramme d'un cube et d'un point de vue, mais il est important de se souvenir que la seule chose que l'ordinateur ait à stocker sont les triplets correspondant à chaque point crucial, rassemblés en groupes plus larges (par exemple un pour chaque face du cube), avec l'information codée pour les diverses propriétés de chaque face (sa couleur, qu'elle soit ou non opaque ou transparente, sa texture, et ainsi de suite). La rotation de l'un des objets et son déplacement dans l'espace virtuel peuvent être aisément calculés, en ajustant simplement toutes les coordonnées *x*,

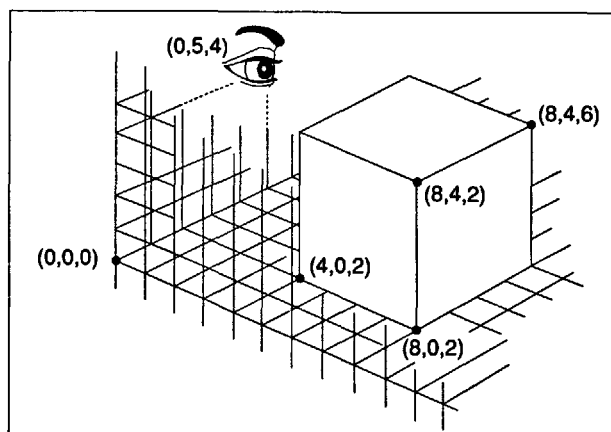
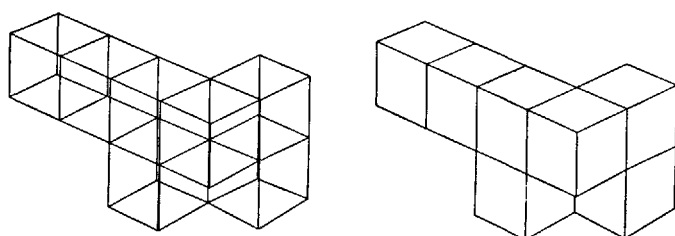


Figure 9.2

y et z de cet objet en quantités constantes – c'est une affaire d'arithmétique élémentaire. C'est ensuite une affaire de géométrie usuelle que de calculer les lignes de vision qui déterminent les parties planes de l'objet qui seraient alors visibles à partir du point de vue virtuel, et exactement comment elles pourraient apparaître. Les calculs sont simples mais laborieux et supposent une « calculation intense », en particulier s'il faut calculer aussi les courbes douces, les ombres, les reflets de lumière et la texture.

Sur des systèmes avancés, on peut calculer différents cadres de manière assez rapide pour créer un mouvement apparent sur l'écran, mais seulement si les représentations restent très schématiques. « La suppression des lignes cachées », le processus informatique qui rend l'image finale opaque aux bons endroits et qui empêche le cube de Shepard de ressembler à un cube de Necker transparent, est elle-même un processus qui prend du temps, et qui limite plus ou moins ce que l'on peut produire en « temps réel ». Pour les transformations merveilleusement détaillées accomplies par les dessins sur ordinateur que nous voyons à la télévision tous les jours, le processus de création des images est bien plus lent, même sur les superordinateurs, et les cadres individuels doivent être emmagasinés pour être réutilisés plus tard à des vitesses plus grandes qui satisfont les réquisits de la direction du mouvement du système visuel humain ².

Ces manipulateurs d'objets virtuels tridimensionnels sont de magnifiques instruments ou jouets nouveaux, et on peut vraiment dire qu'avec eux il y a quelque chose de nouveau sous le soleil,



Avant et après le retrait de la ligne cachée

Figure 9.3

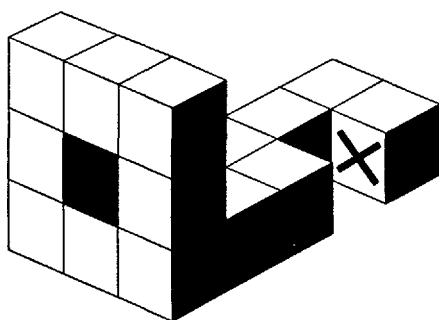


Figure 9.4

qui n'est en rien une copie électronique de quelque chose que nous aurions déjà dans nos têtes. Il est tout à fait certain qu'aucun processus analogue à ces milliards de milliards de calculs géométriques et arithmétiques n'intervient dans nos cerveaux quand nous produisons de l'imagerie mentale, et rien d'autre ne *pourrait* produire les suites animées détaillées qu'ils produisent – pour les raisons que nous avons explorées dans le prélude.

Nous pouvons trouver une compensation dans le fait que les limites de notre cerveau sont réelles en considérant un problème légèrement différent de ceux de Shepard, que l'on pourrait facilement résoudre à l'aide d'un système CAD : le X « rouge » placé sur l'une des faces de cet objet serait-il visible pour quelqu'un qui regarderait à travers le trou carré qui perce la face de devant ?

Notre objet de Shepard avec un X est un objet simple et schématique, et comme la question à laquelle nous voulons répondre est indépendante de la texture, de l'illumination et autres subtilités, un ingénieur pourrait très facilement produire

une rotation animée de cet objet sur le CRT. Il pourrait alors faire tourner l'image dans n'importe quel sens, en déplaçant de-ci de-là le point de vue – et en se contentant de *chercher* un aperçu de rouge à travers le trou. S'il voit du rouge, la réponse est oui, sinon, elle est non.

À présent, pouvez-vous accomplir la même expérience avec l'œil de votre esprit ? Pouvez-vous faire tourner l'objet et regarder par le trou ? Si vous le pouvez, vous pouvez faire quelque chose que je ne peux pas faire, et tous les gens à qui j'ai posé la question sont incapables de le faire en confiance. Même ceux qui ont une réponse à la question sont tout à fait sûrs qu'ils ne l'ont pas fait *seulement* en faisant tourner la figure et en regardant. (Ils disent souvent qu'ils ont d'abord essayé de la faire tourner et de regarder en constatant que cela ne marchait pas ; ils peuvent bien la « faire tourner » mais elle se met alors à se « détacher » quand ils essaient de regarder par le trou. Ils imaginent alors « tirer » des traits à travers le trou sur l'image immobile, pour voir s'ils pourraient dire où les lignes touchent la surface plane à l'arrière.) Étant donné que notre objet de Shepard n'est pas plus complexe que les objets qu'on fait tourner apparemment très bien dans de nombreuses expériences, cela pose une énigme : quel processus peut accomplir si aisément certaines transformations (et extraire des informations du résultat), puis échouer si piteusement sur d'autres opérations qui ne sont apparemment pas plus complexes ? (Si ces opérations *nous semblent* ne pas être plus complexes, c'est que nous les envisageons du mauvais point de vue, car nos échecs montrent qu'elles sont plus complexes.)

Une expérience menée par les psychologues Daniel Reisberg et Deborah Chambers (à paraître) soulève à peu près la même question. On montrait à des sujets prétendant être doués pour les exercices d'imagerie mentale des formes « sans signification »



Figure 9.5

et on leur demandait de les faire tourner mentalement de 90 ou de 180 degrés dans leur esprit, puis de rapporter ce qu'ils « voyaient ». Ils furent surpris de découvrir qu'ils ne pouvaient pas reconnaître, avec l'œil de leur esprit, ce que l'on peut aisément découvrir quand on tourne le livre dans le sens des aiguilles d'une montre de 90 degrés pour regarder ces figures.

Le type de questions auxquelles les ingénieurs répondent en utilisant des systèmes CAD n'est en général pas aussi simple que « Est-ce que le X rouge est visible à travers le trou ? » D'ordinaire ces questions concernent les propriétés spatiales plus complexes des objets construits, telles que « Ce bras de robot à trois articulations sera-t-il capable d'atteindre et d'actionner le bouton de réglage sur le dos du robot sans cogner le bloc d'alimentation ? » ou même des propriétés esthétiques des objets comme « De quoi aura l'air la cage d'escalier de ce hall d'hôtel aux yeux de quelqu'un qui marche dans la rue et regarde à travers les baies vitrées ? » Quand nous essayons de visualiser ce genre de scènes sans nous aider de techniques annexes, nous n'obtenons que les résultats les plus schématiques et les plus incertains ; c'est pourquoi un système CAD peut être considéré comme une prothèse pour l'imagination (Dennett, 1982d, 1990b). Il amplifie beaucoup la capacité d'imagination d'un être humain, mais il dépend du fait que l'utilisateur ait une vision normale – pour regarder sur l'écran de contrôle.

Essayons à présent d'imaginer un dispositif prothésique plus ambitieux : un système CAD pour des ingénieurs aveugles ! Et, pour simplifier, supposons que les questions auxquelles ces ingénieurs aveugles veulent répondre soient ordinaires, géométriques, et non des questions subtiles sur l'esthétique de l'architecture. La sortie devra prendre une forme non visuelle, bien entendu. La forme la plus adaptée à l'utilisateur serait celle de réponses en langage ordinaire (ou en Braille ou en synthèse vocale) à des questions en langage ordinaire. Nous supposerons donc que mis en présence de questions de ce type l'ingénieur aveugle passerait simplement la phrase au système CAD (en termes que ce dernier peut « comprendre », bien entendu) et qu'il attendrait que le système CAD lui fournisse la réponse.

Notre système de Série I CADAVEUGLE (*CADBLIND*) n'est pas élégant, mais il est simple. Il consiste en un système CAD ordinaire, avec son CRT complet, en face duquel est perché un *Vorsetzer*, un système de vision par ordinateur, avec une caméra dirigée

vers le CRT et des doigts robotisés pour manipuler les boutons du système CAD³. À la différence de Shakey, dont le CRT ne servait qu'à des personnes situées en dehors de lui, ce système « regarde » vraiment une image, une image réelle faite de points phosphorescents, qui irradie de la lumière réelle de différentes fréquences sur les transducteurs à lumière sur le dos de la caméra vidéo. Quand on lui soumet notre problème de Shepard du X rouge, le CADAVEUGLE de Série I produit une image où figure un vrai X, visible pour tous, y compris les caméras vidéo du *Vorsetzer*.

Supposons alors, sans difficulté supplémentaire, que le *Vorsetzer* ait résolu suffisamment des problèmes de la vision assistée par ordinateur pour être capable d'extraire l'information recherchée à partir des représentations qui brillent sur l'écran du CRT. (Non, je ne vais pas jusqu'à dire que le *Vorsetzer* est conscient — je veux juste supposer qu'il est assez bon pour faire son travail de manière à pouvoir répondre aux questions que l'ingénieur aveugle lui pose.) Le CADAVEUGLE de Série I fabrique et manipule de vraies images ; il les utilise pour répondre, au bénéfice de l'ingénieur aveugle, à toutes les questions auxquelles pourrait répondre un ingénieur doué d'une vision normale utilisant un système CAD. Si le CADAVEUGLE de Série I est si bon, alors ce sera un jeu d'enfant de construire l'engin de Série II : nous enlevons seulement le CRT et la caméra vidéo qui le regarde, et nous les remplaçons par un simple câble ! À travers ce câble, le système CAD envoie au *Vorsetzer* la *carte à bits*, la gamme de zéros et de uns qui définit l'image sur le CRT. Dans le *Vorsetzer* de Série I, cette carte à bits était péniblement reconstruite à partir des sorties des transducteurs optiques dans la caméra.

Dans la machine de Série II on fait quelques maigres gains en *computation*, en éliminant seulement un peu de matériel inutile. Tout le calcul complexe de lignes de vision, la suppression des lignes cachées et le rendu de texture, d'ombre et de lumière reflétée, qui supposait tant de calculs dans l'engin de Série I, fait toujours partie du processus. Supposez que le *Vorsetzer* dans l'engin de Série II soit mis en œuvre pour produire un jugement sur la profondeur en comparant des gradients de texture, ou en interprétant une ombre. Il aura à *analyser* les trames de bits dans les portions pertinentes de la carte à bits pour parvenir à une discrimination des textures et des ombres.

Cela veut dire que l'engin de Série II est encore une machine

ridiculement inefficace, parce que si l'information selon laquelle une portion particulière de la carte à bits doit représenter une ombre est déjà « connue » par le système CAD (si cela fait partie de la description codée de l'objet à partir duquel le système CAD engendre ses images), et si cela fait partie de ce que le *Vorsetzer* doit déterminer afin de former son jugement sur la profondeur, pourquoi est-ce que le système CAD ne le *dit* pas tout simplement au *Vorsetzer* ? Pourquoi se fatiguer à *rendre* l'ombre au bénéfice des analyseurs de formes du *Vorsetzer*, quand la tâche de transcription de formes et la tâche d'analyse de formes s'annulent réciproquement ?

Par conséquent, notre CADAVEUGLE de Série III s'épargnera d'énormes tâches computationnelles de transcription d'images en prenant la majeure partie de ce qu'il « sait » des objets représentés et en le passant directement au système *Vorsetzer*, en utilisant le format de codes simples pour des propriétés, et en attachant des « labels » à divers « lieux » sur le tableau de la carte à bits, qui se transforme ainsi de pure image en quelque chose comme un diagramme. Certaines propriétés spatiales sont représentées directement – *montrées* – dans l'espace (virtuel) de la carte à bits, mais d'autres sont seulement *énoncées* par des labels ⁴.

Cela devrait nous rappeler ce que j'ai dit au chapitre 4 : le cerveau n'a besoin de faire ses discriminations qu'une seule fois ; un trait identifié n'a pas besoin d'être repassé pour le bénéfice d'un juge en chef dans le Théâtre Cartésien.

Mais à présent nous pouvons voir un autre aspect de l'ingénierie : « l'annulation » ne peut fonctionner que si les systèmes qui ont besoin de communiquer « parlent le même langage ». Que se passe-t-il si le format dans lequel le système CAD « connaît » déjà l'information pertinente – par exemple l'information que quelque chose est une ombre – n'est pas le format dans lequel le *Vorsetzer* peut « utiliser » l'information ⁵ ? Par suite, pour que la communication ait lieu il pourrait être nécessaire de « reculer pour mieux sauter ». Le système pourrait avoir à s'engager dans des interactions prodigues en informations – autant dire interminables – pour pouvoir avoir seulement des interactions. C'est comme si on se mettait à dessiner une carte pour donner des directions à un étranger, quand tout ce qu'il a besoin de savoir – si seulement vous pouviez le lui dire dans sa langue maternelle – est « Tournez à gauche au prochain feu rouge ». Se fatiguer à fabriquer quelque chose comme une image est souvent nécessaire

pratiquement, même quand ce n'est pas nécessaire « en principe ».

Étant donné que les systèmes que nous avons dans nos cerveaux sont les produits de plusieurs histoires empilées de bricolages opportunistes, la longue histoire de la sélection naturelle et la courte histoire de la réorganisation par automanipulation, nous devons nous attendre à rencontrer ce genre de défauts. De plus, il existe d'autres raisons pour transcrire l'information sous une forme imagée (en plus du plaisir que cela procure) qui, si nous avons la chance de tomber dessus, nous frapperont vite comme justifiant la fabrication d'images dans n'importe quel cas. Comme nous l'avons déjà noté, dans les spéculations du chapitre 6 sur l'activité de « se fabriquer des diagrammes », ces transformations de format permettent souvent d'extraire très efficacement une information qui est souvent loin d'être inextricable à partir des données. Les diagrammes reviennent en fait à *représenter* l'information – non pour l'œil de l'esprit, mais pour un mécanisme interne de reconnaissance de formes qui peut aussi accepter des entrées provenant d'un œil normal (« externe » ?). C'est pourquoi les techniques de graphisme assisté par ordinateur sont tellement utiles dans la recherche scientifique, par exemple. Elles permettent la présentation d'énormes quantités de données sous un format qui laisse les superbes capacités de reconnaissance de formes de la vision humaine prendre le relais. Nous faisons des graphes, des cartes et toutes sortes de tracés de couleurs de façon à ce que les régularités et les saillances que nous recherchons « ressortent » tout à coup, grâce à nos systèmes visuels. Les diagrammes ne nous aident pas seulement à voir des trames qui autrement seraient imperceptibles : ils peuvent nous aider à *garder la trace* de ce qui est pertinent, et ils nous *rappellent* de poser les bonnes questions aux bons moments. Le chercheur suédois en IA Lars-Erik Janlert (1985) a soutenu qu'on pouvait aussi utiliser une telle génération d'images et son usage sur ordinateur pour résoudre des problèmes qui seraient autrement insolubles au sujet de ce que l'on pourrait appeler de la gestion d'inférence dans des systèmes qui sont « en principe » de purs moteurs à déduction (pour un autre point de vue sur le même processus, voir Larkin et Simon, 1987).

Bien des chercheurs astucieux connaissent certainement très bien cette technique, et l'un des plus astucieux qui fut jamais, le physicien Richard Feynman, l'a merveilleusement décrite dans

Vous voulez rire Mr Feynman (1985). Dans un chapitre opportunément appelé « Une autre boîte à outils », il nous dit comment il épatait ses condisciples quand il était étudiant à Princeton en « devinant » la vérité ou la fausseté des théorèmes énigmatiques de la topologie, théorèmes qu'il était totalement incapable de démontrer formellement ou même de comprendre complètement :

« J'avais une technique, que j'utilise encore aujourd'hui quand quelqu'un m'explique quelque chose que j'essaie de comprendre : je ne cesse de fabriquer des exemples. Par exemple, les mathématiciens arrivent avec un théorème extraordinaire, et ils sont tout excités. Au fur et à mesure qu'ils me donnent les conditions du théorème, je construis quelque chose qui satisfait toutes les conditions. Vous savez, vous avez un ensemble (une balle) – disjoint (deux balles). Puis, les balles deviennent des couleurs, il leur pousse des poils, ou quoi que ce soit d'autre que je puisse me figurer, sur lequel ils imposent des conditions nouvelles. Finalement, ils énoncent le théorème, qui est une chose incompréhensible au sujet de la balle, mais qui n'est pas vrai de ma balle verte poilue, et par conséquent je dis : " Faux ".

Si c'est vrai, ils deviennent tout excités, et je les laisse faire pendant un moment. Puis, je donne mon contre-exemple.

" Oh ! Nous avons oublié de vous dire que c'est une homomorphie de Hausdorff de Type 2. "

" Eh bien alors, dis-je, ' C'est trivial ! c'est trivial ! ' " À ce moment-là je sais si c'est vrai ou faux, même si je ne sais pas ce que veut dire " homomorphie de Hausdorff " » (p. 85-86).

Ces genres de tactiques « viennent naturellement à l'esprit » dans une certaine mesure, mais il faut les apprendre ou les inventer, et certaines personnes sont meilleures que d'autres pour les réaliser. Ceux chez qui ces aptitudes sont très développées ont des machines virtuelles différentes dans leur cerveau, avec des pouvoirs significativement différents, comparés à ceux qui sont des « visualiseurs » en de rares occasions ou qui sont incapables de l'être. Et les différences apparaissent immédiatement dans leurs mondes hétérophénoménologiques individuels.

On a donc de bonnes raisons de croire, comme l'ont soutenu Kosslyn et d'autres, que les êtres humains mettent en marche leur système de vision non pas seulement en se donnant des images réelles, *externes* (comme sur le CRT du système CAD), mais aussi des images internes qu'ils construisent de façon idiosyn-

crasique ou avec des représentations de données diagrammatiques qui jouent le rôle de matériau de base pour une ou des étapes ultérieures de machinerie de traitement d'information visuelle.

Sur quelles solutions d'ingénierie le cerveau humain est-il tombé et à quels problèmes de communication interne et de manipulation d'information ces solutions correspondent-elles ? Quelles sont leurs forces et leurs faiblesses ⁶ ? Voilà quelques-unes des questions que se pose la recherche sur l'imagerie en psychologie cognitive, et nous devons faire attention quand il s'agit de leur fournir des réponses *a priori* ⁷. Nous pourrions, je suppose, avoir trouvé des systèmes de manipulation d'images de Série II dans nos cerveaux, pleins de points phosphorescents et un œil interne sensible à la lumière. (Pour autant que je puisse voir, il n'est pas *impossible* que des créatures vivant sur une quelconque planète soient dotées de ce genre d'engins.) Et il faut des expériences comme celle de Reisberg et Chambers pour montrer que les raccourcis sur lesquels tombent nos cerveaux empêchent tout à fait que l'on puisse découvrir un système de Série II, avec un format de carte à bits qui ne se sert jamais de raccourcis. (Si nous avions un tel système, il serait facile de résoudre le problème du X rouge, tout comme celui de faire tourner le Texas.)

La phénoménologie nous donne des indices, qui vont dans les deux directions : le caractère « ébauché » des images mentales, qui est « intuitivement évident » dans la phénoménologie de la plupart des sujets, manifeste l'usage de raccourcis par le cerveau, des occasions où le cerveau dit sans montrer. C'est aussi vrai de la *perception* visuelle que de la *visualisation*. Nous avons déjà noté au chapitre 1 combien il est difficile de dessiner une rose qui soit juste devant nos yeux, ou même de copier un dessin. C'est parce que les propriétés purement spatiales que l'on doit identifier ou discriminer pour bien faire bien le dessin ont été normalement laissées de côté au cours du traitement de l'information perceptive, résumées dans des comptes rendus, mais pas reproduites pour des usages ultérieurs. D'un autre côté, l'utilité des images mentales pour nous permettre de « voir la trame », ou pour nous « rappeler » des détails qu'autrement nous oublierions, signale que l'exploitation de la machinerie de reconnaissance des trames visuelles ne peut intervenir que si une partie du cerveau se fatiguait à préparer des versions avec un format spécial de l'information au bénéfice de ces systèmes visuels. Comme nous l'avons déjà vu dans le prélude, les réquisits de la mani-

pulation de l'information contenue dans ces représentations sont fantastiques, et il ne devrait pas être surprenant que nous soyons si peu capables de garder en tête des diagrammes aussi schématiques sous une forme stable.

Voici un test simple pour vous rappeler combien nos capacités sont en fait limitées : *dans l'œil de votre esprit*, remplissez les mots croisés de trois cases sur trois suivants, en écrivant les trois mots suivants dans les colonnes, à partir de la colonne de gauche : TON ÂNE SUE.

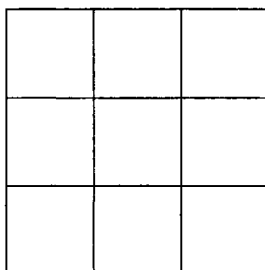


Figure 9.6

Pouvez-vous aisément *déchiffrer* les mots horizontaux ? Dans un véritable diagramme sur le papier, ces mots vous « sauteraient » aux yeux – vous seriez incapable de ne pas les *voir*. Après tout, c'est pour cela qu'on fait des diagrammes : pour présenter des données dans un format qui rende aisée ou inévitable une nouvelle décomposition ou une nouvelle analyse des données. Un tableau de trois cases sur trois n'est pas une structure de données très compliquée, mais ce n'est apparemment pas quelque chose que notre cerveau peut tenir en place de façon assez stable pour que ses systèmes visuels fassent leur travail pour que cela vous « saute aux yeux ». (Si vous voulez essayer deux autres groupes de mots pour les colonnes : CON UNE LUE, et RIZ ILE ZEN.)

Cependant, toutes sortes de variations individuelles peuvent intervenir dans les tactiques employées par différents visualiseurs, et certains peuvent être capables de trouver – ou de développer – des stratégies d'imagerie qui leur permettent de « lire » ces diagrammes. Les calculateurs prodiges sont capables de s'entraîner à multiplier des nombres à dix chiffres de tête, et c'est pourquoi il ne serait pas surprenant que certaines personnes puissent développer des talents prodigieux pour « lire des mots croisés » mentalement. Ces démonstrations informelles nous

donnent certains indices, mais les expériences peuvent définir bien plus précisément les types de mécanismes et de processus que les gens doivent utiliser dans ces actes d'automanipulation. Les données dont on dispose aujourd'hui confirment la thèse selon laquelle nous utilisons des stratégies mixtes, en tirant certains bénéfices de l'analyse visuelle de tableaux, mais aussi en incorporant des étiquettes abrégées, *qui nous disent quelque chose sans nous le montrer*.

Il faut noter cependant, même dans le système CADAVEUGLE de Série II, qui est ultrapictorial et qui incorpore une carte à bits rendant la couleur, les ombres et la texture pixel par pixel, qu'on peut toujours dire – en un sens métaphysique important, comme nous le verrons dans le chapitre suivant – que c'est du « dire » et pas du « montrer ». Considérez le X rouge sur votre figure de Shepard (figure 4). Dans l'engin de Série I il est rendu par du rouge réel – le CRT émet de la lumière, qui doit être transduite par quelque chose dans la caméra de télévision, analogue aux cônes dans vos yeux qui répondent aux différences de fréquence. Quand le *Vorsetzer* fait tourner l'image dans un sens ou dans un autre, à la recherche d'un aperçu de rouge à travers le trou, il attend que ses démons détecteurs de rouge se déclenchent. Dans l'engin de Série II, ce matériel est écarté, et la carte à bits représente la couleur de chaque pixel par un nombre. La nuance de rouge est peut-être le nombre de couleur 37. Le *Vorsetzer* dans l'engin de Série II, quand il fait tourner l'image codée en bits dans un sens ou dans l'autre, cherche à travers le trou un aperçu sur le nombre 37. Ou, en d'autres termes, il demande si l'un des démons à pixels veut bien le lui dire, « Couleur 37 ici. » Tout le rouge est parti – il n'y a que des nombres ici. Finalement, tout le travail dans le système CADAVEUGLE est effectué par des opérations arithmétiques sur des suites de bits, comme nous l'avons vu au niveau inférieur de la vision de Shakey au chapitre 3. Et peu importe si les processus qui aboutissent aux réponses verbales du *Vorsetzer* aux questions ont un caractère quasi pictorial ou imagé : elles ne seront pas produites dans un lieu interne où les propriétés qui ont disparu (les propriétés qui sont simplement « énoncées » dans la carte à bits) sont en quelque façon *restaurées* afin d'être *évaluées* par un juge qui compose les réponses.

Les gens ne sont pas des systèmes CADAVEUGLES. Le fait qu'un système CADAVEUGLE puisse manipuler et inspecter ses « images mentales » sans s'aider d'un Théâtre Cartésien ne prouve pas en

lui-même qu'il n'y ait pas de Théâtre Cartésien dans le cerveau humain, mais cela prouve que nous n'avons pas à postuler un Théâtre Cartésien pour expliquer la capacité qu'ont les humains de résoudre des problèmes « mentalement ». Il y a bien entendu des processus mentaux qui sont fortement *analogues* à l'observation, mais quand nous ramenons la métaphore du CRT à sa plus simple expression, nous écartons les propriétés mêmes qui impliqueraient l'existence d'un Théâtre Cartésien. Il n'est pas besoin d'un lieu ou d'un temps où « tout se rejoint » au bénéfice d'un juge unique et unifié ; les discriminations peuvent s'accomplir de façon distribuée, asynchrone, et sur de multiples niveaux.

Mots, images et pensées

« L'aspect authentiquement " créateur " du langage ne tient pas à sa " capacité générative infinie " mais aux cycles de production et de compréhension rendus possibles par un esprit capable de réfléchir sur les sens multiples qui peuvent être attachés à une énonciation. Ces significations n'ont pas besoin d'être présentes dans la pensée qui a donné lieu à l'énonciation, mais elles deviennent accessibles à la compréhension de soi (ou à l'interprétation profonde de l'énonciation de quelqu'un d'autre) et elles peuvent conduire à une nouvelle pensée qu'on peut réexprimer et réinterpréter, et ainsi de suite. »

H. Stephen Straight, 1976, p. 540.

On demanda un jour à l'économiste britannique John Maynard Keynes s'il pensait en mots ou en images. Il répondit : « Je pense en pensées. » Il avait raison de résister à la suggestion selon laquelle « les choses dans lesquelles nous pensons » sont soit des mots soit des images, car, comme nous l'avons vu, les « images mentales » ne sont pas *simplement* comparables à des images dans la tête, et la pensée « verbale » n'est pas simplement la même chose que le fait de se parler à soi-même. Mais dire que l'on pense en pensées ne nous permet vraiment pas d'avancer. Cela ne fait que repousser la question, parce qu'une pensée n'est que *ce tout qui se passe quand nous pensons* — c'est un point sur lequel il n'y a pas d'accord unanime.

À présent que nous avons envisagé quelle sorte de machinerie pouvait sous-tendre causalement les divers aspects de nos univers hétérophénoménologiques, nous pouvons *commencer* à rendre compte de la phénoménologie de la pensée, en expliquant non seulement les limites et les conditions de la phénoménologie « visuelle » et « verbale », mais aussi en cherchant d'autres sortes de pensées qui échappent à cette dichotomie.

L'un de mes exercices favoris en hétérophénoménologie visuelle est le roman de Vladimir Nabokov, *La Défense Loujine* (1930), qui raconte l'histoire du Grand Maître Loujine, un génie des échecs, victime d'une dépression nerveuse en plein milieu de son match le plus intense. On nous montre trois étapes dans le développement de sa conscience : son esprit quand il est petit garçon (avant sa découverte des échecs à l'âge de dix ans), puis son esprit quand il est saturé d'échecs (jusqu'à sa dépression nerveuse), enfin les restes pitoyables des deux premières périodes après sa dépression nerveuse, quand sa femme l'emprisonne dans un monde où il n'y a plus de jeux d'échecs – où il ne peut plus parler d'échecs, jouer aux échecs, lire des livres d'échecs, et où son esprit revient à une sorte de paranoïa d'enfant gâté, illuminé par des moments repris à l'univers des échecs – des attaques éclairs sur des problèmes d'échecs et des figures empruntées aux rubriques spécialisées des journaux – pour se figer en dernier lieu sur ses obsessions de joueur d'échecs qui culminent dans sa peur de « se mettre mat ». Nous apprenons que Loujine a tellement saturé son esprit de problèmes d'échecs qu'il voit sa propre vie en ces termes. Voici la façon bizarre dont il fait la cour à la personne qui deviendra sa femme :

« Loujine [...] commença, par de prudentes approches, dont le sens ne lui apparaissait pas encore nettement, à faire une originale déclaration d'amour. "Eh bien, racontez-moi encore quelque chose", répétait la jeune fille, bien qu'elle eût remarqué son air silencieux et renfrogné.

Il demeurait assis, appuyé sur sa canne, et considérant un poteau télégraphique, et le tilleul qui se dressait sur une pente éclairée, il se disait qu'on pourrait, d'un mouvement de cavalier, avec celui-ci prendre celui-là ; en même temps il tâchait de se souvenir de quoi il avait parlé » (tr. fr. p. 111-112 ⁸).

« Son épaule appuyée contre la poitrine de Loujine, elle essayait, d'un doigt prudent, de soulever sa paupière, et, sous la légère

pression exercée sur le globe, surgissait une étrange lueur noire, comme si surgissait son cavalier noir qui pourrait tout simplement perdre le pion si Turati l'avancait au septième coup, comme il l'avait fait lors de leur dernière rencontre » (*ibid.*, p. 128).

Et voici un aperçu de son état d'esprit après sa dépression nerveuse :

« Il échoua dans un local rempli de fumée, où étaient attablés des fantômes tonitruants. Dans chaque coin de la salle une attaque se préparait contre lui – et, repoussant plusieurs tables, un seau d'où émergeait un pion en verre au goulot doré, un tambour sur lequel frappait un cheval à crinière, dressé sur ses pattes de derrière, il atteignit une cage vitrée et scintillante qui tournait doucement » (*ibid.*, p. 156).

Ces thèmes sont des « images » à bien des égards, parce que les échecs sont un jeu spatial, où même l'identité des pièces est fixée canoniquement par leurs formes. Mais le pouvoir des échecs sur l'esprit de Loujine ne se réduit pas aux propriétés visuelles ou spatiales des échecs – tout ce que l'on peut représenter par des photographies ou des films d'un échiquier, avec ses pièces en mouvement. Bien sûr ces propriétés visuelles n'ont qu'une saveur superficielle pour son imagination. Bien plus puissante est la *discipline* fournie par les règles et les tactiques du jeu ; c'est la structure abstraite des échecs qui lui est devenue familière jusqu'à l'obsession, et ce sont ses habitudes d'exploration de cette structure qui font aller son esprit d'une « pensée » à une autre « pensée ».

« Loujine s'en voulait de n'avoir rien remarqué à temps, d'avoir laissé l'initiative à l'adversaire, permettant ainsi, dans son aveuglement candide, à la combinaison de se développer. Il décida d'être plus circonspect à l'avenir, de suivre attentivement la suite des coups, s'il devait s'en produire d'autres, et aussi, cela allait de soi, d'entourer sa découverte d'un secret impénétrable, et de se montrer gai, extrêmement gai. Mais à dater de ce jour, il n'eut plus de repos : sans doute eût-il dû inventer une défense contre cette combinaison perfide, pour s'en délivrer ; mais il n'était pas encore possible d'en deviner le but ni la direction fatale » (*ibid.*, p. 236).

Quand vous avez appris pour la première fois à conduire une bicyclette ou une voiture, vous avez rencontré une nouvelle structure de possibilités d'action, avec des contraintes, des signalisations, des ornières, des perspectives, une sorte de labyrinthe comportemental abstrait dans lequel vous avez vite appris à vous

déplacer. Il est devenu bientôt une « seconde nature ». Vous avez rapidement incorporé la structure de ce phénomène externe dans votre structure de contrôle. Au cours de ce processus, vous pouvez avoir eu des périodes d'exploration obsessionnelle, quand vous ne pouviez pas adapter votre esprit aux nouvelles règles. Je me rappelle une brève période de bridgeomanie durant mon adolescence pendant laquelle j'avais des rêves obsessionnels et absurdes de problèmes de bridge. Je reprenais le même coup une centaine de fois, ou je rêvais de « mises » durant des conversations avec mes professeurs ou mes camarades de classe. Mes rêveries hypnagogiques (ces périodes de type hallucinatoire qu'on éprouve occasionnellement quand on s'endort ou quand on s'éveille) étaient remplies de problèmes du genre : « Quelle est la réponse correcte à une mise de trois livres – trois couteaux ou quatre fourchettes ? »

Il est très courant, quand on rencontre une nouvelle structure abstraite dans le monde – une notation musicale, un langage de programmation, le Code civil, un championnat de base-ball – de se retrouver en train de faire péniblement son chemin à travers ses difficultés, en essayant de se créer des sillons mentaux – de creuser le terrain et d'essayer de se familiariser avec la chose. Loujine n'est que le cas extrême : il n'a qu'une seule structure avec laquelle il puisse jouer, et il l'utilise pour tout. À la fin, cette structure finit par dominer toutes les structures d'habitudes de son esprit, canalisant les suites de ses pensées de façon presque aussi rigide que la suite d'instructions à laquelle obéit le programme d'une machine de von Neumann.

Réfléchissez à toutes les structures que vous avez apprises, à l'école et ailleurs : dire l'heure, l'arithmétique, l'argent, les parcours de bus, utiliser le téléphone. Mais de toutes les structures auxquelles nous nous sommes familiarisés durant nos vies, celle qui est certainement la source la plus profonde et la plus puissante de discipline dans notre esprit est certainement notre langue maternelle. On voit souvent mieux en considérant les contrastes ; Oliver Sacks, dans *Seeing Voices* (1989), attire fortement notre attention sur les richesses que le langage apporte à un esprit en rappelant l'appauvrissement terrible subi par l'esprit d'un enfant sourd, si cet enfant ne peut avoir accès rapidement à un langage naturel (le langage Signe, ou un langage signe). Au chapitre 7 nous avons vu comment le vocabulaire même dont nous disposons influence non seulement la façon dont nous parlons aux autres,

mais aussi la façon dont nous nous parlons à nous-mêmes. Au-delà et en plus de cette contribution *lexicale*, il y a la contribution *grammaticale*. Comme le fait remarquer Levelt (1989, sect. 3.6), les structures obligatoires des phrases de notre langage sont autant de guides à notre disposition, qui nous rappellent qu'il nous faut vérifier ceci, faire attention à cela, et qui nous *obligent* à organiser les faits d'une certaine façon. Une partie de cette structure est certainement innée, comme Chomsky et d'autres l'ont soutenu, mais peu importe par où passe la ligne qui divise les structures qui sont génétiquement déposées dans le cerveau et celles qui y entrent sous la forme de mèmes. Ces structures, réelles ou virtuelles, établissent certains des chemins sur lesquels les « pensées » peuvent ensuite voyager.

Le langage infecte et infléchit nos pensées à tous les niveaux. Les mots de nos vocabulaires sont des catalyseurs qui peuvent précipiter des fixations de contenu au moment où une partie du cerveau communique avec une autre. Les structures de la grammaire imposent une discipline à nos habitudes de pensée, en façonnant les techniques par lesquelles nous sondons nos « bases de données », en essayant, comme l'esprit chercheur d'oiseaux de Platon, à faire surgir les bons oiseaux quand nous faisons l'appel. Les structures des histoires que nous apprenons nous servent de guide à un autre niveau, nous poussant à nous poser les questions qui ont le plus de chances d'être pertinentes dans les circonstances du moment.

Tout cela n'a aucun sens tant que nous persistons à considérer l'esprit comme étant idéalement rationnel, et comme quelque chose de transparent et d'unifié. À quoi pourrait bien servir le fait de *se parler à soi-même*, si l'on sait déjà ce qu'on a l'intention de dire ? Mais à partir du moment où l'on voit la possibilité d'une compréhension partielle, d'une rationalité imparfaite, d'une intercommunication problématique entre les diverses parties, nous pouvons voir comment les forces puissantes qu'un langage introduit dans un esprit peuvent être exploitées en marchant dans ses propres pas, quelquefois à son propre bénéfice, quelquefois à son désavantage.

Par exemple :

Tu es magnifique !

Autre exemple :

Tu es pathétique !

Vous savez ce que veulent dire ces phrases. Vous savez aussi que je viens juste de les introduire à l'improviste, dans le but d'avancer une remarque philosophique, et que ce ne sont pas les actes de langage intentionnels de qui que ce soit. Je ne suis certainement en train ni de vous flatter, ni de vous insulter, et il n'y a personne dans les environs. Mais pourriez-vous flatter vous-même, ou vous insulter vous-même, en vous servant de l'une ou l'autre de mes phrases, et en vous la répétant sans cesse, « en accentuant » ? Essayez, si vous osez. Quelque chose se passe. Vous ne croyez pas un instant ce que vous dites (vous vous le dites à vous-même), mais vous découvrez que dire les mots à vous-même suscite des réactions, peut-être une petite rougeur aux oreilles, en même temps que des réponses, des répliques, des dénégations, des images, des souvenirs, des projets. Ces réactions peuvent aller dans un sens ou dans un autre, bien entendu. Dale Carnegie⁹ avait raison de parler du pouvoir de la pensée positive, mais comme la plupart des technologies, penser est plus facile à créer qu'à contrôler. Quand vous vous parlez à vous-même, vous n'avez pas à croire ce que vous dites pour provoquer des réactions. Certaines réactions doivent se produire, et il faut qu'elles soient pertinentes d'une façon ou d'une autre pour la signification des mots avec lesquels vous vous stimulez. À partir du moment où les réactions commencent à se produire, elles peuvent conduire votre esprit vers des lieux où vous finissez par croire ce que vous dites, et c'est pourquoi il vous faut faire attention à ce que vous vous dites à vous-même.

Le philosophe Justin Leiber résume le rôle que le langage joue dans la formation de notre vie mentale :

« Si nous nous comparons à des ordinateurs, nous ne pouvons pas éviter de voir que le langage naturel est notre principal " langage de programmation ". Cela veut dire qu'une vaste portion de notre connaissance et de notre activité est, pour nous, mieux communiquée et comprise dans notre langue naturelle... On pourrait dire que le langage naturel était notre premier grand artefact originel, et puisque, comme nous nous en apercevons de plus en plus, les langages sont des machines, il en est de même pour le langage naturel qui, avec notre cerveau pour le faire tourner, fut

notre première invention de l'ordinateur universel. On pourrait dire cela, si ne s'introduisait pas le soupçon suivant : le langage n'est pas quelque chose que nous avons inventé, mais quelque chose qui est advenu en nous, pas quelque chose que nous avons construit, mais quelque chose en quoi nous nous sommes créés, et recréés » (Leiber, 1991, p. 8).

L'hypothèse que le langage joue ce rôle fondamental dans la pensée pourrait de prime abord sembler être une version de l'hypothèse très discutée selon laquelle il existe un *langage de la pensée*, un véhicule unique dans lequel s'effectuent toutes les opérations cognitives (Fodor, 1975). La différence est importante, cependant. Leiber a raison d'appeler le langage naturel « langage de programmation », mais nous pouvons distinguer les langages de programmation de niveau supérieur (comme Lisp, Prolog et Pascal) du « langage machine » de base ou du langage « assembleur » de base dont ces divers langages sont composés. Les langages de niveau supérieur sont des machines virtuelles, qui créent des structures (temporaires) dans un ordinateur qui le dote d'une certaine structure de points forts et de points faibles. Le prix à payer pour rendre certaines choses « faciles à dire » est qu'on rend certaines autres choses « difficiles à dire » ou même impossibles. Une telle machine virtuelle peut ne structurer qu'une partie de la compétence de l'ordinateur, en laissant intactes d'autres parties de la machinerie. Si l'on garde cette distinction à l'esprit, on peut avancer l'hypothèse que les traits spécifiques d'un langage naturel – le vocabulaire et la grammaire de l'anglais, du chinois ou de l'espagnol, par exemple – conditionnent le cerveau à la façon d'un langage de programmation supérieur. Mais cela ne revient pas à affirmer la thèse douteuse selon laquelle un langage naturel fournit la structure *d'un bout à l'autre*. Bien entendu, Fodor et ceux qui défendent l'idée du langage de la pensée insistent en général sur le fait qu'ils ne parlent pas du niveau auquel les langages *humains* introduisent leurs contraintes. Ils parlent d'un niveau de représentation plus profond, moins accessible. Il est arrivé à Fodor d'exprimer son idée à l'aide d'une amusante confession. Il reconnaissait que quand il faisait le plus gros effort de pensée, la seule sorte d'entités linguistiques dont il était conscient était des séquences du genre : « Vas-y, Jerry, tu peux y arriver ! » Cela peut avoir été ses « pensées » et nous venons de voir qu'elles pouvaient jouer un rôle important pour l'aider à résoudre les problèmes auxquels il se trouvait confronté,

mais elles peuvent difficilement constituer le matériau à partir duquel on peut construire des inférences visuelles, des hypothèses à tester, et les autres transactions postulées du langage de la pensée de base. Keynes avait raison de refuser le choix entre mots ou images ; les véhicules qu'utilise le cerveau n'ont que des rapports lointains avec les véhicules représentationnels de la vie publique.

Rapporter et exprimer

Lentement, mais sûrement nous avons mis en pièces l'idée du Théâtre Cartésien. Nous avons esquissé une solution rivale à l'hypothèse du Signifieur Central au chapitre 7, et nous venons de voir comment résister aux charmes de l'idée d'un CRT interne. Simples coups de sonde, j'en ai peur ; le Théâtre Cartésien tient toujours, et continue d'exercer son attrait tenace sur nos imaginations. Il est temps de changer de tactique et de l'attaquer de l'intérieur, en le faisant exploser en montrant combien cette hypothèse est incohérente dans ses propres termes. Voyons ce qui se passe quand nous suivons la tradition, et prenons les termes de la « psychologie populaire » de tous les jours pour argent comptant. Nous pouvons commencer en reconsidérant certaines des affirmations plausibles qu'Otto a pu avancer au début du chapitre 7 :

Quand *je* parle [disait Otto], je signifie ce que je dis. Ma vie consciente est privée, mais je peux choisir de vous en divulguer certains aspects. Je peux décider de vous dire diverses choses au sujet de mon expérience présente ou passée. Quand je fais cela, je formule des phrases que j'adapte avec soin au sujet dont je veux rendre compte. Je peux aller et venir entre l'expérience et le compte rendu proposé, confrontant les mots à l'expérience pour être sûr que j'ai trouvé *les mots justes* ¹⁰. Ce vin n'a-t-il pas un petit arrière-goût de *pamplemousse*, ou ne me rappelle-t-il pas celui des *framboises* ? Faut-il dire que le ton plus élevé avait l'air *plus fort*, ou bien ne serait-ce pas plutôt qu'il m'a semblé être *plus clair* ou *plus distinct* ? Je me concentre sur mon expérience consciente particulière et je parviens à un jugement sur mots qui feraient le mieux justice à son caractère. Quand j'ai la satisfaction d'avoir forgé un compte rendu précis, je l'exprime. À partir de mon compte rendu introspectif, vous pouvez venir à connaître un trait quelconque de mon expérience consciente.

Une partie de ce message s'accorde en fait fort bien avec le modèle de production du langage que nous avons proposé au chapitre 7. Le processus d'ajustement par va-et-vient des mots au contenu de l'expérience peut être constaté dans le pandémonium qui réunit les démons producteurs de mots aux démons producteurs de contenu. Ce qui manque, bien entendu, c'est le Je intérieur dont le jugement dirige la correspondance entre ces éléments. Mais bien qu'Otto parle de « ce qu'il choisit » et de ce qu'il « juge », l'introspection ne confirme pas vraiment cette idée.

Nous n'avons qu'un accès très limité aux processus par lesquels les mots « nous viennent à l'esprit », même dans les cas où nous parlons délibérément, en répétant silencieusement nos actes de langage avant de les produire. Les candidats au titre de choses à dire viennent simplement d'on ne sait où. Ou bien nous nous retrouvons déjà en train de les dire, ou bien nous nous trouvons en train de les vérifier, tantôt les rejetant, tantôt les modifiant légèrement pour les prononcer ensuite. Mais même ces étapes intermédiaires occasionnelles ne nous donnent pas d'autres indices sur la façon dont nous procédons. Nous nous retrouvons simplement en train d'accepter ou de rejeter tel mot ou tel autre. Si nous avons des raisons de faire ces jugements, nous les contemplons rarement avant l'acte, et il n'y a que rétrospectivement qu'elles nous paraissent évidentes. (Je m'apprêtais à utiliser le mot *minable*, mais je me suis refréné, parce que cela aurait eu l'air trop prétentieux.) Nous n'avons donc réellement aucune visée privilégiée sur les processus qui se passent en nous quand nous allons des pensées aux mots. Ils *pourraient* bien être produits par un Pandémonium, autant que nous puissions en juger.

Mais pourtant [continue Otto] le modèle du Pandémonium laisse de côté un niveau, une étape du processus. Ce qui manque à votre modèle ce n'est pas une projection dans l'« espace phénoménal » d'un Théâtre Cartésien – quelle idée ridicule – mais c'est un autre niveau d'articulation dans la psychologie du locuteur. Il ne suffit pas de dire que les mots se retrouvent mis en chaîne par une sorte de danse interne qui les assemble, pour être ensuite énoncés. S'ils doivent être des *comptes rendus* des états conscients de quelqu'un, ils doivent d'une façon ou d'une autre être fondés sur un acte d'*appréhension* interne. Ce que le modèle du Pandémonium laisse de côté, c'est l'état conscient du sujet qui guide le discours.

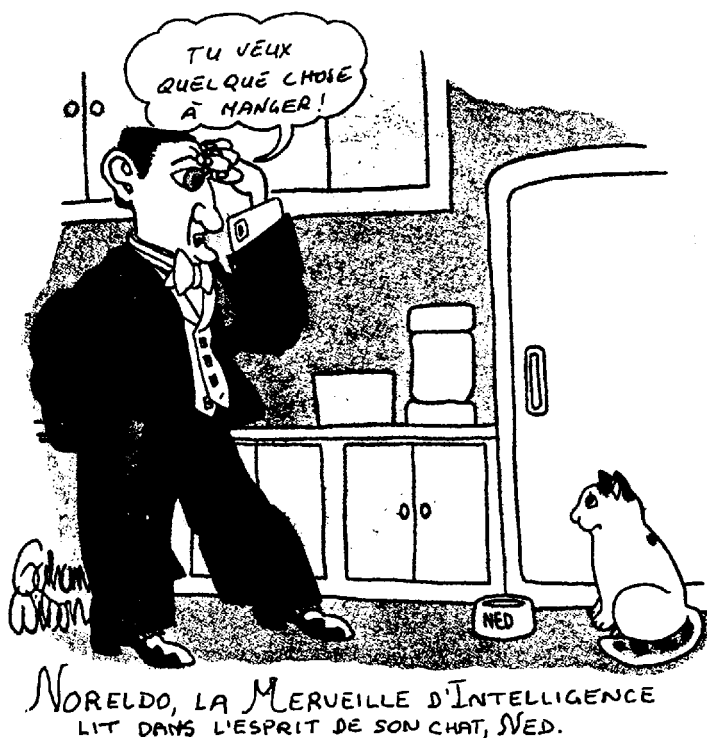
Qu'Otto ait raison ou non, il exprime sans aucun doute la sagesse commune : c'est ainsi que nous concevons ordinairement

notre aptitude à dire aux gens quel est le contenu de nos états conscients. Dans une série d'articles récents, le philosophe David Rosenthal (1986, 1989, 1990a) a analysé ce concept courant de conscience et sa relation à nos concepts de *compte rendu* et d'*expression*. Il découvre certains traits de structure que nous pouvons utiliser avec profit. En premier lieu, nous pouvons utiliser son analyse pour voir, de l'intérieur, en quoi consiste l'image courante, et pourquoi elle est si forte. En second lieu, nous pouvons montrer en quoi elle discrédite l'idée de zombie – sans recourir à des éléments extérieurs. En troisième lieu, nous pouvons retourner l'image courante contre elle-même, et utiliser les difficultés que nous rencontrons pour avancer une image meilleure, qui puisse préserver ce qu'il y a de correct dans la conception traditionnelle mais qui écarte le schéma cartésien.

Que se passe-t-il quand nous parlons ? Au cœur de notre concept courant, il y a un truisme : pour autant que nous ne soyons pas en train de mentir ou non sincères, nous disons ce que nous pensons. En termes plus précis, nous exprimons nos croyances ou pensées. Supposez, par exemple, que vous voyiez le chat en train d'attendre anxieusement à côté du réfrigérateur et que vous disiez : « Le chat veut son dîner. » Cela exprime votre croyance que le chat veut son dîner. En *exprimant* votre croyance, vous *rapportez* ce que vous tenez comme étant un fait au sujet du chat.

Dans ce cas, vous rapportez le désir qu'a le chat d'être nourri. Il est important de noter que vous n'êtes pas en train de *rapporter* votre croyance, ni en train d'*exprimer* le désir du chat. Le chat est en train d'exprimer son désir en se tenant anxieusement devant le réfrigérateur, et vous, remarquant cela, l'utilisez comme la base – la donnée – de votre compte rendu. Il y a bien des manières d'exprimer un état mental (comme un désir) mais il n'y a qu'une seule façon d'en rapporter un, qui consiste à énoncer un acte de langage (oral, écrit ou signalé par un autre moyen).

L'une des façons les plus intéressantes d'*exprimer* un état mental consiste à *rapporter* un autre état mental. Dans l'exemple, vous rapportez le désir du chat, exprimant par là votre propre croyance au sujet du désir du chat. Votre comportement confirme non seulement que le chat a le désir mais aussi que vous croyez que le chat a le désir. Vous pourriez, cependant, avoir fourni la preuve de votre croyance d'une autre façon – peut-être en vous levant sans un mot de votre chaise et en préparant son dîner



Dessin de Gahan Wilson, © 1990 The New Yorker Magazine, Inc.

Figure 9.7

au chat. Cela aurait *exprimé* la même croyance sans *rapporter* quoi que ce soit. Ou encore vous auriez pu simplement vous asseoir dans votre fauteuil et avoir levé les yeux au ciel, en exprimant *de façon non intentionnelle* votre exaspération face au désir du chat alors même que vous veniez juste de vous asseoir confortablement dans votre fauteuil. Exprimer un état mental, délibérément ou pas, c'est seulement faire quelque chose qui confirme bien, ou qui rend manifeste, cet état aux yeux d'un autre observateur – un déchiffreur d'esprits, si vous voulez. Rapporter un état mental, par contre, est une activité plus sophistiquée, toujours intentionnelle, et qui implique le langage.

Voici, par conséquent, un indice important quant à la source du modèle du Théâtre Cartésien : notre psychologie populaire de tous les jours traite l'action de rapporter nos propres états mentaux sur le modèle de nos actions de rapporter des événements

du monde extérieur. Quand vous rapportez que le chat désire son dîner, vous vous fondez sur l'observation du chat. Votre compte rendu exprime votre croyance que le chat désire son dîner, qui est une croyance portant sur le désir du chat. Appelons les croyances qui portent sur des croyances, les désirs qui portent sur des désirs, des espoirs portant sur des craintes, et ainsi de suite, les états mentaux *de second-ordre*. Si je (1) *crois* que vous (2) *pensez* que je (3) *veux* avoir une tasse de café, ma croyance est une croyance *du troisième-ordre* (sur l'importance des états mentaux d'ordre supérieur, voir ma *Stratégie de l'interprète*, 1987a). Il n'y a aucun doute qu'il y a des différences nettes, marquées par ces distinctions usuelles quand on les applique de façon non réfléchie – quand x croit que y est dans un certain état mental et que $x \neq y$. Il y a toute la différence du monde entre le cas où le chat désire être nourri et que vous le savez, et le cas où le chat désire être nourri et que vous ne le savez pas. Mais que se passe-t-il dans les cas réfléchis, où $x = y$? La psychologie populaire traite ces cas comme identiques.

Supposez que je rapporte que *je* désire être nourri. Selon le modèle usuel, je dois être en train d'exprimer une croyance *du second-ordre* au sujet de *mon* désir. Quand je rapporte mon désir, j'exprime une croyance *du second-ordre* – ma croyance sur mon désir. Que se passe-t-il si je *rapporte* cette croyance *du second-ordre*, en disant « Je crois que je désire être nourri » ? Ce compte rendu doit exprimer une croyance *du troisième-ordre* – ma croyance que je crois bien en fait croire que je désire être nourri. Et ainsi de suite. Nos concepts usuels de ce qu'implique le fait de parler sincèrement engendrent de cette façon quantité d'états mentaux supposés distincts : mon désir est distinct de ma croyance que j'ai le désir, qui est distinct de ma croyance que j'ai la croyance que j'ai le désir, et ainsi de suite.

La psychologie populaire fait aussi d'autres distinctions. Comme le fait remarquer Rosenthal (avec beaucoup d'autres), elle distingue les *croyances*, qui sont les états dispositionnels sous-jacents, des *pensées*, qui sont des états occurrents ou épisodiques, des événements passagers. Votre *croyance que les chiens sont des animaux* a persisté continûment comme un état de votre esprit pendant des années, mais en venant d'attirer votre attention sur elle j'ai provoqué en vous une *pensée* – la pensée que les chiens sont des animaux, épisode qui sans aucun doute ne serait pas intervenu en vous si je ne l'avais pas provoqué.

Il s'ensuit, bien entendu, qu'il peut y avoir des pensées du premier-ordre et du second-ordre et d'ordre supérieur – des pensées sur des pensées (sur des pensées...). Voici, par conséquent, le point crucial : quand j'exprime une croyance – comme ma croyance que je désire être nourri, je n'exprime pas la croyance d'ordre supérieur directement ; ce qui se passe, c'est que ma croyance sous-jacente produit une pensée épisodique, la *pensée* d'ordre supérieur que je désire être nourri, et j'exprime *cette* pensée (si je choisis de le faire). Rosenthal soutient que tout cela est impliqué dans le modèle qu'a le sens commun de ce qui se passe quand on *dit ce que l'on pense*.

Puisque l'un des traits distinctifs des états de conscience humaine est qu'on peut les rapporter (sauf cas d'aphasie, de paralysie, ou les cas où l'on est ligoté et bâillonné par exemple), il s'ensuit, selon l'analyse de Rosenthal, que « les états conscients doivent être accompagnés par des pensées d'ordre supérieur appropriées, alors que les états non conscients ne peuvent pas être accompagnés de ce type de pensées » (1990b, p. 16). La pensée d'ordre supérieur en question doit bien entendu porter sur l'état qu'elle accompagne : elle doit être une pensée que l'on est dans l'état d'ordre inférieur (ou qu'on vient d'être dans cet état – le temps s'écoule). Il semble que cela va produire une régression à l'infini d'états conscients d'ordre supérieur ou de pensées, mais Rosenthal soutient que la psychologie populaire nous permet de faire une inversion frappante : *la pensée du second-ordre n'a pas elle-même à être consciente pour que son objet du premier-ordre soit conscient*. On peut *exprimer* une pensée sans être conscient de cette pensée, et par conséquent, on peut exprimer une pensée du *second-ordre* sans être conscient de cette pensée – tout ce dont on a besoin d'être conscient est son objet, à savoir la pensée de premier-ordre que l'on rapporte.

Cela peut paraître surprenant au premier abord, mais à la réflexion nous pouvons reconnaître que c'est un fait familier vu dans une perspective différente : vous n'êtes pas attentif à la pensée que vous exprimez, mais à l'objet sur lequel *porte* cette pensée. Rosenthal soutient aussi que bien que certaines pensées du second-ordre *soient* effectivement conscientes – en vertu de pensées du troisième-ordre qui portent sur elles – celles-ci sont relativement rares. Ce sont les pensées explicitement rétrospectives que nous ne rapporterions (même à nous-mêmes) que dans un état d'hyper-conscience de soi. Si je vous dis « J'ai mal », je

rapporte un état conscient, ma douleur, et j'exprime une croyance du second-ordre – ma croyance que j'ai mal. Si, prenant un ton philosophique, je dis « Je pense (ou je suis sûr, ou je crois) que j'ai mal », je rapporte par là une pensée du second-ordre, en exprimant une pensée du troisième-ordre. Néanmoins, ordinairement, je n'aurais pas une telle pensée du troisième-ordre, et par conséquent je ne serais pas conscient d'une telle pensée du second-ordre ; je l'exprimerais en disant « J'ai mal », mais je ne serais pas ordinairement conscient de cette pensée.

L'idée qu'il puisse y avoir des pensées inconscientes d'ordre supérieur peut sembler loufoque ou paradoxale de prime abord, mais la catégorie d'épisodes en question n'est pas vraiment sujette à controverse, même si on n'utilise pas habituellement le terme de « pensée » pour les désigner. Rosenthal utilise « pensée » au sens technique – en suivant en gros la pratique qui était celle de Descartes, en fait – pour désigner *tous* les états épisodiques doués de contenus, et pas seulement les épisodes que nous appellerions ordinairement des pensées. Ainsi un élancement de douleur ou un regard furtif jeté sur un bas féminin compteraient comme des pensées pour Descartes et Rosenthal. À la différence de Descartes, cependant, Rosenthal insiste sur l'existence de pensées inconscientes.

Les pensées inconscientes sont, par exemple, des événements perceptifs inconscients ou des activations épisodiques de croyances, qui surviennent naturellement – qui *doivent* survenir – durant le contrôle du comportement normal. Supposez que vous renversiez votre tasse de café sur votre bureau. Brusquement, vous sautez de votre chaise, et évitez de justesse le café qui s'égoutte du bord de la table. Vous n'étiez pas conscient de penser que le plateau de la table n'absorberait pas le café, ou que le café, qui est un liquide obéissant à la loi de la pesanteur, se répandrait par-dessus le bord de la table, mais ce genre de pensées inconscientes doivent s'être produites – parce que si la tasse avait contenu du sel de table, ou si le bureau avait été couvert d'un torchon, vous n'auriez pas sauté en arrière. Parmi toutes vos croyances – sur le café, la démocratie, le base-ball, sur le prix du thé en Chine – celle-ci et quelques autres ont immédiatement été pertinentes dans les circonstances où vous vous trouviez. Si je les cite pour expliquer pourquoi vous avez sauté en arrière, c'est qu'elles doivent avoir été saisies ou activées momentanément ou tirées d'une façon ou d'une autre de votre esprit pour pouvoir

contribuer à votre comportement. Mais tout cela s'est bien entendu produit inconsciemment. Ces épisodes inconscients seraient des exemples de ce que Rosenthal appelle des pensées inconscientes. (Nous avons déjà rencontré les pensées inconscientes dans certains cas antérieurs : par exemple les perceptions inconscientes de vibrations dans les doigts qui vous permettent d'identifier consciemment les textures des choses touchées avec une baguette, ou le souvenir inconscient de la femme aux lunettes, qui conduisait à l'expérience trompeuse de la femme qui courait.)

Rosenthal fait remarquer qu'en trouvant une façon de définir la conscience en termes d'états mentaux *inconscients* (les pensées d'ordre supérieur accompagnant les autres), il a découvert une façon d'asseoir à l'intérieur de la *psychologie populaire* les fondements d'une théorie non circulaire, non mystérieuse, de la conscience (1990b). Ce qui distingue un état conscient d'un état non conscient, soutient-il, n'est pas une quelconque propriété intrinsèque inexplicable, mais la propriété banale d'avoir une pensée de second-ordre qui accompagne l'état en question et porte sur lui. (Voir Harnad, 1982, pour une stratégie similaire, avec quelques variations intéressantes.) C'est bon signe pour la psychologie populaire : elle ne se vautre pas simplement dans le mystère ; elle a les ressources, bien dégagées par Rosenthal, d'articuler une analyse de sa catégorie favorite, la conscience, en termes de ses catégories subsidiaires et moins problématiques. Bonne affaire, car si nous optons pour cette analyse, nous pourrions aussi démonter la distinction tranchée entre les êtres conscients et les zombies.

Les zombies, les zimbos et l'illusion de l'utilisateur

« L'esprit est une structure perçue par un esprit.
C'est peut-être circulaire, mais ce n'est ni un cercle
vicieux ni un paradoxe. »

Douglas Hofstadter, 1981, p. 200.

Les zombies des philosophes, vous vous en souvenez, semblent accomplir des actes de langage, semblent rapporter leurs états de conscience, faire des introspections. Mais ils ne sont pas réel-

lement conscients, en dépit du fait qu'on ne peut distinguer leur comportement de celui d'une personne consciente. Ils peuvent avoir des états internes dotés de contenu fonctionnel (le type de contenu que les fonctionnalistes peuvent assigner à la machinerie interne des robots), mais ce sont des états inconscients. Shakey, tel que nous l'avons imaginé, est un zombie paradigmatique. Quand il « rapporte » un état interne, ce n'est pas un état conscient qui est rapporté, parce que Shakey n'a pas d'états conscients, mais un état inconscient qui le conduit causalement vers un autre état inconscient qui dirige le processus d'engendrement et qui exécute un soi-disant acte de langage composé de formules « mises en boîte ». (Nous avons laissé Otto dire cela tout du long.)

Shakey n'a pas décidé *d'abord* ce qu'il devait rapporter, après avoir observé quelque chose qui se passait en lui, pour ensuite se figurer la façon dont il pourrait l'*exprimer* ; Shakey s'est juste trouvé avoir certaines choses à dire. Shakey n'avait pas d'accès aux raisons pour lesquelles il voulait dire qu'il formait des dessins autour des limites ombrées de ses images mentales. Il était seulement fait comme ça. La thèse centrale du chapitre 7, cependant, était que, contrairement aux apparences, c'est vrai également de vous. Vous n'avez pas d'accès spécial aux raisons pour lesquelles vous voulez dire ce que vous découvrez avoir envie de dire ; vous êtes seulement fait comme cela. Mais, à la différence de Shakey, vous êtes constamment en train de vous reconstruire vous-même, découvrant de nouvelles choses que vous avez envie de dire après avoir réfléchi sur ce que vous vous êtes découvert en train d'avoir envie de dire. Et ainsi de suite.

Mais un Shakey plus recherché ne pourrait-il pas le faire lui aussi ? Shakey était un zombie particulièrement grossier, mais nous pouvons à présent imaginer un zombie plus réaliste et plus complet, qui contrôle ses propres activités, y compris même ses propres activités internes, dans une spirale ascendante indéfinie de réflexivité. J'appellerai une telle entité réflexive un *zimbo*. Un zimbo est un zombie qui, du fait de son autocontrôle, a des états informationnels internes (mais inconscients) d'ordre supérieur, qui portent sur ses états informationnels d'ordre inférieur. (Le fait qu'un zimbo soit considéré comme un robot ou comme un être humain – ou Martien ne change rien à cette expérience de pensée.) Ceux qui croient que le concept de zombie est cohérent doivent sûrement accepter la possibilité d'un tel zimbo. Un zimbo est simplement un zombie qui est comportementalement

complexe, grâce à un système de contrôle qui permet une représentation de soi récursive.

Considérons maintenant la question de savoir comment un zimbo pourrait réaliser le test de Turing, le test opératoire fameux proposé par Alan Turing (1950) pour détecter la pensée chez un ordinateur. Un ordinateur peut penser, a affirmé Turing, s'il peut régulièrement battre un adversaire humain au « jeu de l'imitation » : les deux protagonistes du jeu sont cachés aux yeux d'un juge humain, mais ils sont capables de communiquer avec le juge en tapant des messages avec des terminaux d'ordinateurs. Le protagoniste humain essaie simplement de convaincre le juge qu'il (ou elle) est humain(e), tandis que le protagoniste ordinateur en fait autant – c'est-à-dire qu'il essaie de convaincre le juge que *lui*, ordinateur, est humain. Si le juge ne parvient pas régulièrement à détecter l'ordinateur, l'ordinateur est considéré comme un être pensant. Selon Turing, ce test serait propre à éliminer toute contestation ; il était évident à ses yeux que ce test était si facile à passer que n'importe quel ordinateur qui pourrait gagner devrait être considéré par n'importe qui comme un être *fantastiquement* bon pour produire des pensées. Il pensait qu'il avait placé la barre assez haut pour que cela satisfasse n'importe quel sceptique. Il s'était trompé. Beaucoup de gens ont soutenu que « passer le test de Turing » ne serait pas une preuve suffisante d'intelligence, et certainement pas une preuve de conscience. (Voir Hofstadter 1981b, Dennett, 1985a, et French, 1991, pour des analyses des forces et des faiblesses du test de Turing et de ses critiques.)

Or les chances d'un zimbo dans le test de Turing devraient être aussi bonnes que celles de n'importe quelle personne consciente, puisque les protagonistes ne montrent au juge rien d'autre que leur comportement et seulement leur comportement verbal (taper à la machine). Supposez, par conséquent, que vous soyez le juge dans le test de Turing, et que les actes de langage d'un zimbo (apparent) vous aient convaincu qu'il est conscient. Ces actes de langage apparents ne devraient pas vous avoir convaincu – par hypothèse, puisqu'il n'est qu'un zimbo et que les zimbos ne sont pas conscients. Devrait-il s'être convaincu lui-même, cependant ? Quand un zimbo fait un compte rendu, exprimant son propre état inconscient du second-ordre, il n'y a rien qui l'empêche de réfléchir (inconsciemment) sur cet état de choses précis. En fait, s'il doit être convaincant, il devra être capable

de répondre de façon appropriée à ses propres « affirmations » adressées à vous (ou d'en prendre connaissance).

Supposez, par exemple, que le zimbo soit un Shakey plus sophistiqué, et que vous, le juge, veniez juste de lui demander de résoudre un problème mentalement et de vous expliquer comment il y est parvenu. Il réfléchit sur sa propre affirmation qu'il vient juste de résoudre le problème de former une ligne sur une image mentale. Il « saurait » que c'est ce qu'il a voulu dire, et s'il réfléchissait un peu plus, il viendrait à « savoir » qu'il ne savait pas pourquoi c'était ce qu'il voulait dire. Plus vous lui en demandez sur ce qu'il savait et ne savait pas sur ce qu'il était en train de faire, plus il deviendrait capable de réfléchir. À présent ce que nous avons réussi à imaginer, semble-t-il, c'est un être inconscient qui a néanmoins la capacité d'avoir des pensées d'ordre supérieur. Mais selon Rosenthal, quand un état mental est accompagné d'une pensée d'ordre supérieur, consciente ou inconsciente, portant sur l'existence même de cette pensée, cela garantit *ipso facto* que l'état mental est un état conscient ! Notre expérience de pensée discrédite-t-elle l'analyse de Rosenthal, ou discrédite-t-elle la définition du zimbo ?

Nous pouvons voir tout de suite qu'à tout le moins le zimbo croirait (inconsciemment) qu'il a divers états mentaux – précisément les états mentaux qu'il était en position de rapporter si on lui posait des questions. *Il* (le zimbo) penserait qu'il est conscient, tout en ne l'étant pas ! Toute entité qui passerait le test de Turing aurait le sentiment (erroné ?) d'être consciente. En d'autres termes, elle serait victime d'une illusion (*cf.* Harnad, 1982). Quelle sorte d'illusion ? Une illusion de l'utilisateur, bien entendu. Elle serait la « victime » de l'illusion bénigne de l'utilisateur sur sa propre machine virtuelle !

N'est-ce pas un jeu de miroirs, une sorte de tour de passe-passe illégitime de philosophe ? Comment pourrait-il y avoir une illusion de l'Utilisateur sans un Théâtre Cartésien dans lequel l'illusion se trouve perpétrée ? Il semble que je sois en danger de me laisser avoir par mes propres métaphores. Le problème est que l'illusion de l'Utilisateur d'une machine virtuelle *est* en fait réalisée par une présentation des données sur une sorte de théâtre, et qu'il faille aussi un auditoire indépendant et extérieur, à savoir l'Utilisateur, pour qui on monte le spectacle. Je suis en ce moment même en train d'utiliser un ordinateur, tapant ces mots sur un « fichier » avec l'assistance discrète d'un programme

de traitement de texte. Quand j'entre en interaction avec l'ordinateur, j'ai un accès limité aux événements qui se produisent en lui. Grâce aux schèmes de présentation conçus par les programmeurs, je suis invité à une métaphore audiovisuelle, à un drame interactif qui se joue sur la scène du clavier, de la souris et de l'écran. Moi, l'Utilisateur, suis sujet à une série d'illusions bénignes : j'ai l'air d'être capable de mouvoir le curseur (un serviteur puissant et fidèle) à l'endroit même dans l'ordinateur où je veux avoir mon dossier, et une fois que j'ai vu que le curseur est arrivé « là », en pressant une touche je lui fais sortir mon fichier, en le répandant sur un long ruban qui l'enroule face à une fenêtre (l'écran) à mes ordres. Je peux faire advenir toutes sortes de choses à l'intérieur de l'ordinateur en tapant diverses commandes, en pressant divers boutons, et je n'ai pas besoin de connaître les détails ; je maintiens le contrôle en m'appuyant sur ma compréhension des métaphores audiovisuelles précises que me fournit l'illusion de l'Utilisateur.

Pour la plupart des utilisateurs d'ordinateurs, seules ces métaphores permettent d'évaluer ce qui se passe à l'intérieur. C'est l'une des raisons pour lesquelles il est si tentant de comparer la conscience à une machine virtuelle, parce qu'il nous a toujours semblé que notre accès à ce qui se passe à l'intérieur de nos cerveaux était limité : *nous* n'avons pas à savoir comment la machinerie en coulisse de nos cerveaux fait ses tours de magie ; *nous* sommes familiarisés avec ses opérations seulement sous l'habillage qu'elles prennent pour nous sous la forme des métaphores interactives de la phénoménologie. Mais si, quand nous nous autorisons cette analogie tentante, nous maintenons la séparation « évidente » entre la Présentation d'un côté et l'Évaluation du spectacle par l'Utilisateur de l'autre, il semble que nous revenions précisément au Théâtre Cartésien. Comment pourrait-il y avoir une illusion de l'Utilisateur si cette séparation n'existait pas ?

Il ne pourrait pas y en avoir une ; l'utilisateur qui fournit la perspective à partir de laquelle la machine virtuelle devient « visible » doit être une sorte d'observateur extérieur – un *Vorsetzer*. Et on pourrait penser de prime abord que le concept d'un tel observateur doit être le concept d'un observateur conscient. Mais nous avons déjà vu que ce n'est pas le cas. Le *Vorsetzer* qui s'asseyait en face du système CAD dans le système CADAVEUGLE de Série I n'était pas conscient, mais il était cependant exactement

aussi limité dans son accès aux mécanismes internes du CAD que n'importe quel utilisateur conscient, et à partir du moment où nous écartons les accessoires gratuits de l'écran et de la caméra, la Présentation et l'Évaluation de l'Utilisateur s'évaporent, et sont remplacées, comme cela nous arrive si souvent dans notre théorie, par une quantité de transactions plus modestes. L'« observateur extérieur » peut être graduellement incorporé dans le système, laissant derrière quelques traces fossiles : des petits bouts d'« interface » dont les divers formats continuent à conditionner les sortes de questions auxquelles on peut répondre, ce qui en retour conditionne les contenus exprimables¹¹. Il n'y a pas besoin qu'il existe un lieu *unique* où se produit la Présentation¹². Et comme nous le suggère l'analyse de Rosenthal, même notre concept ordinaire de conscience, tel qu'il est ancré dans les intuitions du sens commun ou de la psychologie populaire, peut tolérer le caractère inconscient des états d'ordre supérieur dont la présence dans le système rend compte du caractère conscient de certains de ses états.

Le processus de réflexion inconsciente est-il, par conséquent, l'une des voies par lesquelles un zombie pourrait se transformer en zimbo et *par là* devenir conscient ? Si oui, alors les zombies doivent être, après tout, conscients. Tous les zombies sont capables d'énoncer des « actes de langage » convaincants (rappelez-vous que l'on ne peut les distinguer de nos meilleurs amis) ; cette capacité serait magique si les structures de contrôle ou les processus qui en sont causalement responsables dans le cerveau du zombie (ou ordinateur, ou quoi que ce soit) n'étaient pas capables de réfléchir sur les actes et sur leurs contenus (apparents ou fonctionnels). Un zombie doit commencer sa carrière dans un état de non-communication et de non-réflexion, et par conséquent être vraiment un zombie, un être inconscient. Mais dès qu'il aurait commencé à « communiquer » avec d'autres et avec lui-même, il deviendrait équipé des sortes d'états précisément qui, selon l'analyse de Rosenthal, sont suffisants pour qu'il y ait conscience.

Si, d'un autre côté, on rejette l'analyse donnée par Rosenthal de la conscience en termes de pensées d'ordre supérieur, alors les zombies devront se prêter à d'autres expériences de pensée. J'offre cette parabole des zimbos sur le mode de la plaisanterie, parce que je ne pense pas *non plus* que le concept de zombie ni les catégories de pensées d'ordre supérieur de la psychologie popu-

laire pourront survivre sinon comme reliques d'une croyance démodée. Mais Rosenthal nous a rendu service en exposant la logique de ces concepts de tous les jours, et grâce à la conception claire que nous en avons à présent, nous pouvons voir comment on pourrait les remplacer par de meilleurs.

Les problèmes de la psychologie populaire

Rosenthal a découvert que la psychologie populaire postule une hiérarchie indéfinie de pensées d'ordre supérieur, conçues comme des épisodes saillants, indépendants, dotés de contenus et se produisant en temps réel dans le cerveau. Que devient cette conception, si nous cherchons à la confirmer. Y a-t-il des états et des événements distincts de ce genre dans le cerveau ? Si nous sommes généreux, la réponse doit être oui. Il ne fait pas de doute qu'il existe des différences psychologiques familières qu'on peut décrire – et qui sont en général décrites – dans ces termes.

Dorothée réalisa soudain qu'elle voulait partir – et qu'elle avait voulu partir depuis un bon bout de temps.

Ici, il semble que Dorothée ait acquis la croyance du second-ordre – en ayant une pensée du second-ordre – portant sur son désir quelque temps après que le désir eut pris effet. Il y a de nombreux cas de ce type dans la vie de tous les jours : « Et c'est alors qu'il s'aperçut qu'il était en train de regarder droit dans la direction du bouton de manchette qui lui manquait. » « Il l'aime, mais il ne le réalise pas encore. » On peut difficilement nier que ces phrases ordinaires fassent allusion à des transitions authentiques d'un « état d'esprit » à un autre. Intuitivement, comme Rosenthal le fait remarquer, la transition repose sur *la conscience que l'on prend* de l'état du premier-ordre. Quand Freud, à partir de ces cas ordinaires, postule un vaste domaine d'états mentaux inconscients, ce sont précisément les états dont les sujets ne croient pas qu'ils les aient. Ces gens sont dans des états d'esprit dont ils ne s'aperçoivent pas encore – par l'intermédiaire de pensées d'ordre supérieur – qu'ils les ont.

Cette forme de description de ces différences nous est familière, mais savoir si elle est entièrement adéquate est une autre affaire. Ces états sont tous des transitions menant à un état de meilleure information (pour le dire de la façon la plus neutre possible), et

être bien informé de cette façon est bien entendu une condition nécessaire pour *rapporter* (par opposition à simplement *exprimer*) l'« état d'esprit » antérieur. Or il serait imprudent de formuler cela de la façon suivante : *afin de rapporter* un état ou événement mental, il vous faut avoir une pensée d'ordre supérieur que vous exprimez. L'image que cela suggère est qu'on *observe* d'abord (avec un organe sensoriel interne quelconque) l'état ou événement mental, produisant *par là même* un état de *croissance* dont l'origine est marquée par une *pensée*, qui se trouve alors exprimée. La chaîne causale, comme on l'a vu, mime la chaîne causale par laquelle on rapporte des événements extérieurs ordinaires. Vous observez d'abord l'événement avec l'aide de vos organes sensoriels, produisant en vous une croissance, puis une pensée, que vous exprimez dans un compte rendu.

Cette pensée supposée d'ordre supérieur est, je pense, le « niveau d'articulation supplémentaire » qu'Otto pouvait discerner dans sa propre psychologie ; c'est la pensée que les mots d'Otto *expriment* quand il *rapporte* sa propre expérience consciente. Mais selon le modèle de production du discours que nous avons esquissé au chapitre 7, le modèle d'Otto inverse tout simplement l'ordre causal. Nous ne formons pas *d'abord* un état d'ordre supérieur d'auto-observation, qui crée une pensée de second-ordre, en sorte que l'on puisse ensuite rapporter la pensée d'ordre inférieur en exprimant la pensée d'ordre supérieur. Ce qui se passe plutôt c'est que l'état du second-ordre (l'état de meilleure information) vient à être *créé* par le processus même de formation du compte rendu. Nous n'appréhendons pas *d'abord* notre expérience dans le Théâtre Cartésien, puis, sur la base de cette connaissance acquise, acquérons l'aptitude à former des comptes rendus que nous exprimons ; *le fait même que nous soyons capables de dire* quel effet cela nous fait *est la base même* de nos « croyances d'ordre supérieur » ¹³.

De prime abord, un processus de conception d'acte de langage par Pandémonium semble mal expliquer le phénomène parce qu'il semble écarter l'observateur/décideur central dont la pensée sera finalement exprimée. Mais c'est une force de ce modèle, pas une faiblesse. L'émergence de l'expression est précisément ce qui crée ou fixe le contenu de la pensée d'ordre supérieur exprimée. Il n'est pas nécessaire qu'il y ait une « pensée épisodique » *additionnelle*. L'état d'ordre supérieur dépend littéralement de – dépend causalement de – l'expression de l'acte de langage. Mais il ne dépend pas nécessairement de l'expression publique d'un acte de langage

explicite. Au chapitre 6, nous avons vu comment le besoin qu'a l'organisme d'une communication interne de l'information sans cesse meilleure pourrait avoir conduit à la création d'habitudes d'automanipulation qui pourraient prendre la place des processus évolutifs plus laborieux de création d'un œil intérieur, d'un organe interne réel qui pourrait contrôler le cerveau. La seule façon dont un cerveau humain pourrait se placer dans quelque chose qui ressemble à un état de croyance d'ordre supérieur, à mon sens, c'est de s'engager dans un processus qui ressemblerait à celui de se rapporter à lui-même des états d'ordre supérieur.

Il nous faut rompre avec l'habitude de postuler sans cesse de nouveaux observateurs centraux. En guise de transition, nous pouvons reconsidérer le processus comme n'impliquant pas de la connaissance par observation, mais sur le modèle de l'*oui-dire*. Je crois que *p* parce que l'on m'a dit que *p* de source sûre. Qui me l'a dit ? Moi-même, ou en tout cas un – ou plusieurs – de mes « agents ». Ce n'est pas une façon de parler si étrange que cela. Après tout, nous parlons du *témoignage* de nos sens, et cette métaphore suggère que nos sens n'apportent pas de preuves matérielles devant le « tribunal » pour nous les *montrer*, mais plutôt pour nous *dire* des choses. En nous appuyant sur cette métaphore (jusqu'à ce que nous nous habituions aux complexités d'une solution meilleure), nous pourrions proposer un slogan :

Si je pouvais me parler à moi-même, je n'aurais aucun moyen de savoir ce que j'étais en train de penser.

Ce n'est pas encore tout à fait ça, à plusieurs égards. Tout d'abord, il y a une différence – déjà commentée – entre une entité qui se « parle à elle-même » et divers systèmes « qui se parlent entre eux ». On négociera la transition entre ces deux idées au chapitre 12, sur le moi. En second lieu, comme nous l'avons vu, il est exagéré d'insister sur l'expression linguistique : il y a d'autres stratégies d'automanipulation et d'expression de soi qui ne sont pas verbales.

J'ai peut-être l'air de proposer un mauvais marchandage : abandonner la relative netteté et la clarté du modèle usuel de la psychologie populaire, avec sa hiérarchie d'observations internes, en faveur d'une option encore obscure. Mais la clarté du modèle traditionnel est une illusion, pour des raisons que nous avons déjà mentionnées au chapitre 4, quand nous explorions l'idée étrange d'*apparence réelle*. À présent nous pouvons diagnostiquer

les problèmes plus précisément. Otto est le représentant de la psychologie populaire, et si nous le laissons continuer il s'em mêlera bientôt dans ses propres nœuds. La conception d'Otto, qui se rattache obstinément aux catégories psychologiques populaires et les étend « complètement », engendre une explosion d'« états représentationnels » distincts, dont les relations mutuelles engendrent des difficultés artificielles. Otto continue :

Mon compte rendu public d'un état conscient, si je devais en choisir un, pourrait contenir une erreur. Je pourrais faire un lapsus, ou me tromper sur le sens d'un mot et vous induire en erreur sans le vouloir. Toute erreur d'expression que je ne pourrais pas saisir serait susceptible de créer en vous une croyance erronée quant aux faits — quant à ce que j'éprouve *réellement*. Et le simple fait que je n'aie pas pu saisir une erreur ne signifierait pas qu'il n'y avait pas d'erreur. *D'un côté*, il y a la vérité quant à ce que j'éprouve, et *de l'autre côté* il y a ce que je finis par dire sur ce que j'éprouve (si je choisis de le faire). Bien que je tende à rapporter ce qui se produit en moi de façon très fiable, il peut toujours se produire des erreurs.

C'est l'une des situations où il ne suffit pas d'être deux. Car, comme nous l'a montré Rosenthal, en plus de « ce que j'éprouve » et « ce que je finis par dire » il semble qu'il doit y avoir un tiers fait intermédiaire : ma croyance quant à ce que j'éprouve ¹⁴. Car quand je dis sincèrement ce que je dis, en voulant dire ce que cela signifie, j'exprime l'une de mes croyances — ma croyance sur ce que j'éprouve. En fait, il y a un quatrième fait intermédiaire : ma pensée épisodique sur ce que j'éprouve.

Ma croyance sur ce que j'éprouve pourrait-elle être erronée ? Ou bien pourrais-je *seulement penser* ce que j'éprouve ? Ou, en d'autres termes, ne pourrait-il pas me *sembler* seulement que c'était mon expérience du moment ? Otto voulait une séparation plus nette, mais à présent nous risquons d'avoir une séparation plus forte : entre l'expérience subjective et la croyance qui porte sur elle, entre la croyance et la pensée épisodique qu'elle introduit dans le processus d'expression publique, et entre cette pensée et son expression finale. Et comme les balais de l'apprenti sorcier qui se multiplient, il y a encore plus de séparations en perspective à partir du moment où nous acceptons celles-ci. Supposez que j'aie mon expérience subjective (c'est une chose) et qu'elle soit au fondement de ma croyance que j'ai cette même expérience (c'est une deuxième chose) qui en retour introduit une pensée

associée (c'est une troisième chose) qui m'incite ensuite à avoir une intention communicative de l'exprimer (une quatrième chose), qui produit finalement une expression effective (une cinquième chose). L'erreur ne peut-elle se glisser à chaque étape ? Ne pourrait-il se faire que je croie une certaine proposition, mais qu'en raison d'une mauvaise transition entre états, je *viennne à penser* une autre proposition ? (Si vous pouvez « parler de travers » ne pouvez-vous pas « penser de travers » ?) Ne serait-il pas possible de former l'intention d'exprimer une proposition assez différente de celle que vous êtes en train de penser ? Et un souvenir défectueux dans le système de communication ne pourrait-il pas vous conduire à vous apprêter à exprimer un message préverbal, et à finir avec un autre message préverbal qui joue le rôle de critère permettant de corriger les erreurs ? Entre deux choses distinctes, il y a une place logique pour l'erreur, et au fur et à mesure que nous multiplions les états individuels avec des contenus définis, nous découvrons – ou créons – des sources multiples d'erreur.

On peut être très tenté de couper ce sac de nœuds en déclarant que *ma pensée (ou croyance) quant à ce qui me semble être le cas* est tout simplement la même chose que *ce qu'est effectivement mon expérience*. En d'autres termes, on peut être tenté d'insister sur le fait qu'il n'y a pas de place logique pour de l'erreur, parce que ce sont une seule et même chose. Cette thèse a des propriétés intéressantes. Elle arrête l'explosion menaçante à l'étape initiale – ce qui est en général le bon endroit pour arrêter une explosion ou une régression à l'infini – et elle est intuitivement attrayante, car on peut la formuler rhétoriquement : quel sens cela aurait-il de dire que quelque chose m'a seulement semblé m'avoir semblé m'avoir semblé (m'avoir semblé...) être un cheval ?

Mais nous devons ici faire attention et marcher à pas de loup autour des squelettes des théories philosophiques défuntes (y compris certaines des miennes – cf. Dennett, 1969, 1978c, 1979a). On dirait que nous pouvons nous en tenir aux bonnes vieilles catégories de la psychologie populaire, les croyances, les pensées, les expériences et leurs semblables, et éviter les difficultés de la connaissance de soi en mêlant simplement les cas réflexifs des ordres supérieurs et ceux des ordres inférieurs : en déclarant que quand je crois que je crois que *p*, par exemple, il s'ensuit logiquement que je crois que *p*, et dans le même esprit, quand je pense que j'ai mal, il s'ensuit logiquement que j'ai mal, et ainsi de suite.

Si c'était le cas, quand j'ai exprimé une croyance du second ordre en rapportant une croyance du premier ordre, par exemple, je ne me suis occupé que d'un seul état, et le fait qu'en rapportant une chose j'en aie exprimé « une autre » serait dû à une simple distinction verbale, comme le fait que Dupond ait décidé d'épouser sa fiancée et se soit retrouvé en train d'épouser sa femme.

Mais on n'arrive pas à nos fins en pratiquant cette assimilation. Pour s'en rendre compte, considérons à nouveau le rôle de la mémoire, telle que la psychologie populaire la conçoit. Même s'il est intuitivement plausible que vous ne puissiez pas vous tromper quant à ce que vous *ressentez maintenant*, il n'est pas intuitivement plausible que vous ne puissiez pas vous tromper sur ce que vous *ressentiez alors*. Si l'expérience que vous êtes en train de rapporter est une expérience passée, votre mémoire – sur laquelle vous vous reposez dans votre compte rendu – pourrait être contaminée par l'erreur. Peut-être votre expérience était-elle en fait ceci, mais vous vous rappelez à tort qu'elle était cela. Il peut certainement vous sembler *maintenant* qu'il vous avait semblé *alors* que c'était un cheval – même si *en fait* il vous avait semblé *alors* que c'était une vache. La possibilité logique de faux souvenirs s'ouvre quelle que soit la durée de l'intervalle temporel qui s'écoule entre l'expérience effective et les souvenirs ultérieurs – c'est justement ce qui permettait aux théories orwelliennes d'intervenir. Mais, comme nous l'avons vu au chapitre 4, l'erreur qui se glisse dans les croyances ultérieures à la suite des altérations mémorielles orwelliennes est indistinguable – de l'extérieur comme de l'intérieur – de l'erreur qui se glisse dans l'expérience initiale à la suite des inventions d'illusions staliniennes. C'est pourquoi même si nous pouvions maintenir que vous avez un accès « direct » et « immédiat » à votre *jugement* présent (votre pensée du second ordre sur ce que les choses vous semblent être maintenant), vous ne serez pas par là même capable d'exclure la possibilité que ce soit un jugement *erroné* quant à ce que les choses vous ont semblé être un moment auparavant.

Si nous individualisons les états (croyances, états de conscience, états d'intentions de communication, etc.) par leurs contenus – ce qui est la forme d'individuation canonique de la psychologie populaire –, nous finissons par postuler des différences qui sont systématiquement impossibles à découvrir, de l'extérieur comme de l'intérieur, et en cours de route, nous perdons l'intimité subjective ou l'incorrigibilité qui est supposée être la pierre de touche

de la conscience. Nous avons déjà vu des cas de ce genre dans la discussion du chapitre 4 et des modèles des phénomènes temporels orwelliens et staliniens. Et on n'aura pas la solution si l'on s'en tient à l'une ou l'autre des doctrines que la psychologie populaire met à notre disposition : il faudra abandonner ce trait de la psychologie populaire.

Nous remplaçons la division entre *états* discrets dotés de contenu – croyances, métacroyances, et ainsi de suite – par un *processus*, qui sert, à la longue, à assurer un ajustement entre les événements internes porteurs d'information d'une certaine entité, et la capacité qu'a cette entité d'*exprimer* par le discours (une partie de) l'information venant de ces événements. C'est ce que les états d'ordre supérieur étaient supposés assurer (par Otto), mais ils ne donnaient pas le bon découpage ni les bonnes articulations de la nature. En fait, ils impliquaient l'existence d'articulations que l'on ne peut discerner dans la nature.

Et pourtant, ces artefacts de la psychologie populaire continuent à exister, comme habitants des mondes hétérophénoménologiques des sujets dont les conceptions du monde sont de fait façonnées par leurs schèmes conceptuels. Pour l'exprimer par une tautologie, puisqu'il semble vraiment aux gens qu'ils ont à la fois ces croyances au sujet de leurs expériences et (en plus) les expériences elles-mêmes, ces expériences et les croyances-sur-les-expériences font toutes deux partie de la façon dont les choses leur apparaissent. C'est pourquoi il nous faut expliquer *ce fait-là* – pas le fait que nos esprits *sont* organisés en hiérarchies d'« états » représentationnels d'ordre supérieur de croyances, de métacroyances, et ainsi de suite, mais le fait que nos esprits *tendent à nous apparaître* comme organisés de cette façon.

Je me suis risqué à avancer deux raisons pour lesquelles cette idée nous paraît si attrayante. Tout d'abord, nous persistons à avoir l'habitude de postuler l'existence d'un processus séparé d'observation (à présent d'observation *interne*) intervenant entre les circonstances à propos desquelles nous pouvons rapporter quelque chose et le compte rendu que nous pouvons produire – et ce faisant, nous oublions qu'à un moment quelconque cette régression à l'infini d'observateurs intérieurs devra être arrêtée par un processus qui unifie les contenus et leur expression verbale sans qu'intervienne un évaluateur des contenus intermédiaire. Ensuite, les communications internes ainsi produites ont en fait pour effet d'organiser notre esprit en un ensemble indéfini de

systèmes puissants réflexifs ou capables d'autocontrôle. On a souvent dit que ces pouvoirs de réflexion étaient au cœur de la conscience. Cette idée est juste. Nous pouvons utiliser le modèle trop simple de la psychologie populaire comme une sorte de béquille pour guider notre imagination quand nous essayons de comprendre les systèmes d'autocontrôle, mais quand nous utilisons ce modèle, nous risquons de retomber dans le matérialisme cartésien. Il faudra bien que nous commencions à apprendre à nous en passer. Au chapitre suivant, nous ferons quelques pas de plus en ce sens.

NOTES

1. Cette simplification utile est l'une des nombreuses pistes que les chercheurs ont explorées par la suite, et on dispose aujourd'hui de données considérables pour affirmer qu'il y a des effets d'« inertie » et de « moment » dans les transformations d'images. Voir Freyd (1989).

2. L'animation impressionnante, mais nettement tremblante, du programme populaire « Simulateur de Vol » (*Flight Simulator*) sur IBM-PC montre les limites de l'animation en temps réel de scènes tridimensionnelles complexes sur un petit ordinateur.

3. J'appelle ce dispositif un *Vorsetzer* parce qu'il me rappelle la merveilleuse machine à jouer du piano allemande, qui consistait en une unité séparée avec quatre-vingt-huit « doigts » qui pouvait « s'asseoir » en face d'un piano ordinaire, appuyer sur les touches et les pédales du dehors, tout comme un pianiste humain. (Il est important de reconnaître que ce dispositif est un *Vorsetzer* – une machine assise devant – mais pas un *Vorsitzer* – un président ou chef de département.)

4. À partir du moment où nous avons les labels, nous pouvons parler de *n'importe quelles* propriétés de l'objet, pas seulement des propriétés spatiales ou visibles – comme dans les blagues des vieux livres d'images : « Voilà mon directeur. Colore-le antipathique. »

5. Voir Kosslyn (1980) pour une discussion du format. Jackendoff (1989) fournit une analyse similaire de ce qu'il appelle la forme des structures informationnelles.

6. Kosslyn (1980) non seulement défend de manière détaillée son ensemble particulier de réponses à ces questions (à cette époque), mais il fournit aussi un excellent résumé des autres travaux expérimentaux et théoriques sur l'imagerie. Un bon résumé des travaux de la décennie suivante se trouve dans Farah (1988) et dans Finke, Pinker et Farah (1989).

7. Comme le dit Marvin Minsky : « Il n'y a rien de particulièrement problématique dans l'idée de sentir des événements dans le cerveau. Les agents sont des agents – et il est aussi facile pour un agent d'être câblé pour détecter un événement cérébral causé par le cerveau, que de détecter un événement cérébral causé par les monde » (1985, p. 151).

8. Trad. fr. Genia et René Cannac, revue par Bernard Kreise, Gallimard, Folio, Paris, 1991. Le texte donné par Dennett de la traduction anglaise est légèrement différent (NdT).

9. L'Émile Coué américain (NdT).

10. En français dans le texte (NdT).

11. Au chapitre 7 (p. 277), j'ai demandé « excités à la surface de quoi ? » et

j'ai promis de répondre à la question plus tard. Voici ma réponse. La surface (métaphorique) est déterminée par le format des interactions entre les parties.

12. Il est intéressant de comparer différentes traces de cette idée de l'Utilisateur dans le cerveau qu'on trouve chez des penseurs très différents. Voici ce que dit Minsky (1985) : « Pour exagérer un peu, ce que nous appelons "la conscience" consiste en un petit peu plus que des listes de menus qui apparaissent, de temps en temps, sur les écrans mentaux que d'autres systèmes utilisent (p. 57)... Divisez le cerveau en deux parties, A et B. Connectez les entrées et sorties du cerveau A au monde réel, afin qu'il puisse sentir ce qui s'y passe. Mais ne connectez pas le cerveau B au monde extérieur ; au lieu de cela, rattachez-le au cerveau A de façon à ce que le cerveau A soit le monde du B (p. 59) ! Minsky évite sagement de s'aventurer à proposer des lignes de division anatomiques, mais d'autres gens sont prêts à s'aventurer un peu sur ce terrain. Quand Kosslyn émit pour la première fois ses spéculations sur la conscience comme machine virtuelle, il fut enclin à localiser l'utilisateur dans les lobes frontaux (voir aussi Kosslyn, 1980, p. 21), et plus récemment Edelman a suivi son propre chemin argumentatif vers la même conclusion, en termes de "mémoire soi/non-soi dominé par des valeurs", qu'il localise dans les lobes frontaux et à laquelle il assigne la tâche d'interpréter les productions du reste du cerveau » (Edelman, 1989, p. 102 *sq.*).

13. Cela ressemble d'assez près à un thème central dans l'œuvre du second Wittgenstein, mais Wittgenstein refusa de développer une explication positive ou un modèle de la relation entre ce que nous disons et ce dont nous parlons quand nous rapportons (en apparence) nos états mentaux. La philosophe Elisabeth Anscombe, dans son œuvre classique – mais frustrante par son obscurité – *Intention* (1957), essaya de combler cette lacune laissée par Wittgenstein, en soutenant qu'il était faux de dire que nous *savons* ce que sont nos intentions ; c'est plutôt que nous *pouvons dire* ce que sont nos intentions. Elle essaya aussi de définir une catégorie de choses que nous pouvons *savoir sans observation*. On trouve une discussion défectueuse de ces thèses dans mon livre *Content and Consciousness* (1969), chapitre 8 et 9. J'ai toujours pensé qu'il y avait quelque chose de correct, d'important et d'original dans ces passages. J'y suis revenu dans « Toward a Cognitive Theory of Consciousness » (1978), repris dans *Brainstorms*, en particulier sections IV et V (p. 164-171 dans *Brainstorms*). Cette section représente ma tentative présente pour éclairer ces idées, et je diverge ici fortement des deux autres versions.

14. Dans *Brainstorms*, j'ai exploité ce trait de la psychologie populaire dans ma discussion des « β -multiplicités de croyances phénoménologiques » (1978a, p. 177 *sq.*).

Chapitre 10

Le programme de protection du témoin démantelé

Résumé

Dans la première partie, nous avons passé en revue les problèmes et mis en place quelques présupposés et principes méthodologiques. Dans la seconde partie, nous avons esquissé un modèle de la conscience, le modèle des Versions Multiples, et nous avons entrepris de montrer pourquoi on devait le préférer au modèle traditionnel du Théâtre Cartésien. Bien que l'idée du Théâtre Cartésien, une fois rendue explicite, révèle ses défauts de façon tout à fait radicale – il n'existe pas de matérialistes cartésiens déclarés –, les hypothèses d'arrière-plan et les habitudes de pensée qu'elle a promues continuent à motiver des objections et à fausser nos « intuitions ». À présent, dans cette troisième partie, nous sommes en train d'analyser les implications de notre modèle rival, en répondant à un ensemble d'objections pressantes. Certaines de ces objections trahissent une allégeance persistante – même si elle est inavouée – à notre cher vieux Théâtre Cartésien.

« Mais où la compréhension se produit-elle ? » Cette question s'est tenue au centre de la controverse depuis le XVII^e siècle. Descartes s'est heurté à un mur de scepticisme lorsqu'il a insisté (à bon droit) sur le fait que les mécanismes qui agissent dans le cerveau pourraient rendre compte d'au moins une partie de la compréhension. Par exemple, Antoine Arnauld, dans ses objections aux *Méditations*, a noté que « cela semble incroyable d'abord, qu'il puisse se faire, sans le ministère d'aucune âme, que la lumière qui réfléchit du corps du loup dans les yeux de la brebis remue les petits filets des nerfs optiques, et qu'en vertu de ce

mouvement, qui va jusqu'au cerveau, les esprits animaux soient répandus dans les nerfs en la manière qu'il est nécessaire pour faire que la brebis prenne la fuite » (1641, *Pléiade*, p. 428). Descartes répondit que ce n'était pas plus incroyable que la capacité humaine de projeter ses bras en avant quand on tombe, qui est aussi une réaction mécanique, sans le bénéfice d'une « âme ». Cette idée d'une interprétation « mécanique » dans le cerveau est l'idée centrale de toute théorie matérialiste de l'esprit, mais elle menace une intuition profondément enracinée : notre sentiment que, pour qu'une compréhension *réelle* se produise, il faut qu'il y ait *quelqu'un là-dedans* qui valide les actes, pour *être le témoin* des événements dont l'occurrence constitue la compréhension. (Le philosophe John Searle exploite cette intuition dans son expérience de pensée de la Chambre Chinoise, que nous envisagerons au chapitre 13.)

Descartes était mécaniste *par excellence* à propos des phénomènes de la nature ; mais quand il s'agissait des phénomènes de l'esprit humain il déclarait forfait. Il soutenait qu'en plus d'assurer des interprétations mécaniques, le cerveau fournissait des matériaux à une arène centrale – ce que nous avons appelé le Théâtre Cartésien – où, chez les êtres humains, l'âme peut être un Témoin et parvenir à ses propres jugements. Les témoins ont besoin de matériaux bruts sur lesquels fonder leurs jugements. Ces matériaux bruts, qu'on les appelle des « données sensorielles », des « sensations », des « données brutes » ou des « propriétés phénoménales de l'expérience », sont des accessoires sans lesquels la notion même de Témoin n'a pas de sens. Ces accessoires, mis en place par diverses illusions, entourent l'idée d'un Témoin Central doué d'une barrière quasiment infranchissable d'intuitions. Le but de ce chapitre est de renverser cette barrière.

La vision aveugle : le fait d'un zombie partiel ?

De tous les accidents terribles dont peuvent être victimes les gens, il y en a une petite fraction que rachète le fait qu'ils peuvent révéler aux chercheurs quelque chose des mystères de la nature. C'est vrai en particulier des lésions cérébrales produites par un traumatisme (blessures par balle, accidents de la circulation, et autres choses du même genre), une tumeur ou un coup¹. Les

formes d'incapacité ou de capacités réduites fournissent quelquefois des données substantielles – et même étonnantes – sur la façon dont le cerveau produit l'esprit. L'un des plus surprenants de ces phénomènes, comme le suggère son nom, est la vision aveugle. Ce phénomène semble de prime abord avoir été taillé sur mesure pour les expériences de pensée des philosophes. C'est une affection qui transforme une personne normale, consciente, en partie en un zombie, en un automate non conscient de certains stimuli, bien qu'elle soit normalement consciente du reste. Il n'est donc pas surprenant que les philosophes aient élevé la vision aveugle à une sorte de statut mythique comme exemple autour duquel construire des arguments. Pourtant, comme nous allons le voir, la vision aveugle n'apporte aucun argument en faveur de l'existence des zombies ; au contraire, elle a pour effet de miner ce concept même.

Dans la vision humaine normale, les signaux qui entrent par les yeux voyagent à travers les nerfs optiques par l'intermédiaire de diverses étapes vers le *cortex occipital* ou cortex visuel, la partie du cerveau qui se trouve tout à fait à l'arrière du crâne au-dessus du cervelet. L'information qui porte sur le champ visuel gauche (les moitiés gauches du champ à partir de chaque œil) est répandue sur le cortex visuel droit, et le champ visuel droit sur le cortex gauche. Il peut arriver qu'un accident vasculaire (par exemple l'éclatement d'un vaisseau sanguin) détruise une portion du cortex occipital, créant un tache aveugle ou *scotome*, un trou relativement grand dans le monde visuel perçu, du côté opposé à celui de la lésion.

Dans le cas extrême où à la fois les cortex gauche et droit ont été détruits, la personne devient complètement aveugle. Plus souvent, l'ensemble du cortex visuel sur un côté du cerveau est détruit par un accident vasculaire, conduisant à la perte de la partie opposée du champ visuel ; une perte du cortex visuel gauche produirait une *hémianopie droite*, un aveuglement complet dans le champ droit.

Quel effet cela fait-il d'avoir un scotome ? Il pourrait sembler que nous le sachions déjà, parce que nous avons tous des taches aveugles dans nos champs visuels correspondant à l'emplacement de nos rétines où il n'y a pas de bâton ni de cônes parce que c'est là que le nerf optique touche le globe oculaire. Une tache aveugle normale, ou un disque optique, n'est pas petite : elle rend blanc un cercle d'environ six degrés d'angle visuel de dia-

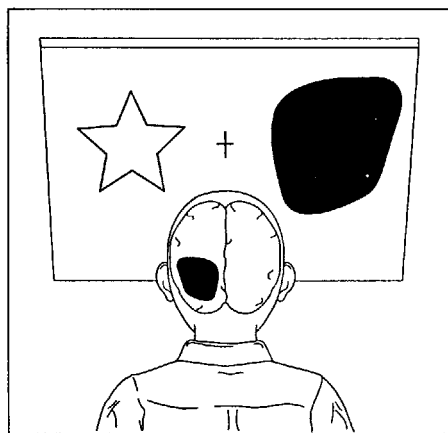


Figure 10.1

mètre. Fermez un œil et regardez la croix, en tenant la page à environ vingt-cinq centimètres de vos yeux. L'un des disques « aveugles » disparaîtra. Fermez l'autre œil, et le disque opposé disparaîtra. (Il peut être nécessaire d'ajuster un peu la page pour faire apparaître l'effet. Continuez à regarder droit dans la croix.) Pourquoi ne remarquez-vous pas habituellement cette lacune dans votre champ visuel ? En partie parce que vous avez deux yeux, et qu'un œil couvre l'autre ; leurs taches aveugles ne se chevauchent pas. Mais même avec un œil fermé vous ne remarquerez pas la tache aveugle dans les conditions normales. Pourquoi ? Parce que votre cerveau n'ayant jamais eu à s'occuper d'entrées venues de cette aire de votre rétine, il n'a pas consacré de ressources à s'en occuper. Il n'y a pas d'homuncules qui soient responsables de la réception de rapports venus de cette aire, en sorte que quand aucun rapport ne parvient au cerveau, il n'y a personne pour se plaindre. Une absence d'information n'est pas la même chose qu'une information au sujet d'une absence. Pour que vous voyiez un trou, quelque chose dans votre cerveau doit répondre à un contraste : *ou bien* entre le bord intérieur et le bord extérieur – et votre cerveau n'a rien dans sa machinerie pour répondre à cet endroit – *ou bien* entre l'avant et l'après : à présent vous voyez le disque, et ensuite vous ne le voyez plus. (C'est comme ça que le disque noir de la figure 10.2 vous alerte sur la présence de la tache aveugle.)

Comme nos taches aveugles normales, les scotomes ont des



Figure 10.2

localisations bien définies et des limites précises qu'un expérimentateur peut repérer facilement, en utilisant un stimulus comme un point lumineux qu'on peut mouvoir autour du champ visuel du sujet. On demande au sujet de rapporter le moment où il ne fait plus l'expérience du point lumineux – ce qui n'est qu'une variante de l'expérience que vous avez conduite sur vous-même pour découvrir votre tache aveugle. Les comptes rendus du sujet peuvent alors être mis en corrélation avec des cartes de la lésion sur le cortex, produits par des scannings CT (*computer aided tomography*, tomographie assistée par ordinateur) et MTI (*magnetic resonance imaging*, images par résonance magnétique) sur le cerveau. Un scotome diffère d'un point de vue important d'une tache aveugle normale : habituellement, le sujet le remarque. Ce n'est pas seulement parce qu'il est plus vaste que la tache aveugle normale. Parce qu'il est causé par la perte de cellules dans le cortex visuel, ces autres cellules qui « reçoivent » l'information dans le cortex qui « font attention » aussi à l'information venue de certaines régions de la rétine, leur absence est remarquée. Les attentes du cerveau sont troublées ; quelque chose qui devrait être là manque à l'appel, il y a un appétit épistémique qui n'est pas satisfait. Le sujet est donc habituellement conscient du scotome, mais c'est pour lui un manque, pas une aire positive de noir, comme vous pourriez la relever si quelqu'un collait un cercle de papier noir sur le parebrise de votre voiture.

Étant donné que les voies visuelles normales dans le cerveau ont été interrompues ou coupées, on devrait s'attendre à ce que les gens affectés d'un scotome soient complètement incapables de glaner la moindre information de la vision des choses qui se produisent dans leurs champs visuels aveugles. Ils sont, après tout, aveugles. Et c'est précisément ce qu'ils disent : ils ne font l'expérience de rien qui soit visuel à l'intérieur des limites de leur scotome – pas de traits lumineux, de contours, de couleurs, de scintillements ni de petites étoiles. Rien. C'est cela être aveugle. Mais certaines personnes affectées d'un scotome manifestent une

capacité étonnante : en dépit de leur totale absence d'expérience visuelle consciente dans l'aire aveugle, elles peuvent quelquefois « deviner » avec une précision remarquable si une lumière vient ou non de s'allumer dans leur champ visuel, et même si on leur a montré un cercle ou un carré. C'est le phénomène de ce que l'on appelle la vision aveugle (Weiskrantz, 1986, 1988, 1990). L'explication de ce phénomène est encore sujette à controverse, mais aucun chercheur ne pense qu'il y ait quoi que ce soit de « paranormal » là-dedans. Il y a au moins encore dix chemins différents entre la rétine et le reste du cerveau. Par conséquent, même si le cortex occipital est détruit, il trouve encore beaucoup de canaux de communication par-dessus lesquels l'information venue des rétines – qui restent parfaitement normales – peut atteindre les autres surfaces cérébrales. On a réalisé aujourd'hui beaucoup d'expériences sur des sujets victimes de vision aveugle, et il ne fait pas de doute qu'ils peuvent réussir bien mieux qu'en s'en remettant au hasard (ils réussissent quelquefois dans 100 % des cas dans certaines conditions) quand il s'agit de deviner diverses formes simples, les directions des mouvements, la présence ou l'absence de la lumière. Aucune personne victime de vision aveugle n'a pour le moment révélé de capacité à discriminer des couleurs dans le champ aveugle, mais des travaux récents de Storig et de Cowey montrent que c'est peut-être possible.

Que se passe-t-il dans la vision aveugle ? Est-ce, comme l'ont soutenu certains philosophes et psychologues, de la perception visuelle sans conscience, du type de celle que pourrait manifester un simple automate ? Refute-t-elle (ou menace-t-elle sérieusement) les théories fonctionnalistes de l'esprit en révélant un cas dans lequel toutes les *fonctions* de la vision sont encore présentes, mais où tout le bon jus de la *conscience* s'est vidé ? Ce phénomène ne prouve rien de tel. Dans leur impatience d'accrocher la vision aveugle à leurs wagons idéologiques, les philosophes ont quelquefois négligé certains faits élémentaires au sujet des phénomènes de vision aveugle et des conditions expérimentales dans lesquelles ils se produisent.

Comme les « anomalies temporelles » analysées aux chapitres 4 et 5, les phénomènes de vision aveugle n'apparaissent que quand nous traitons les sujets du point de vue de l'hétérophénoménologie. On ne pourrait pas conduire les expériences si les expérimentateurs ne pouvaient pas donner des instructions verbales aux sujets (et être sûrs d'avoir été compris). Et les réponses des

sujets ne révèlent un phénomène étonnant *que* quand elles sont interprétées comme des actes de langage. C'est presque trop évident, et c'est pourquoi je dois enfoncer le clou.

L'interprétation de la vision aveugle est à bien des égards sujette à controverse, mais elle est remarquablement non problématique à un égard : tout le monde est d'accord sur le fait que le sujet victime de vision aveugle vient d'une façon ou d'une autre à recevoir une certaine information concernant un événement quelconque dans le monde par l'intermédiaire de ses yeux (c'est la partie qui porte sur la « vision »), en dépit du fait qu'il n'a pas d'expérience visuelle normale de l'événement en question (c'est la partie qui porte sur son caractère « aveugle »). En bref, la vision aveugle implique (1) une réception de l'information visuelle qui est (2) néanmoins inconsciente. La démonstration de (1) est simple : le sujet réussit sur des tests qui sondent l'existence de cette information mieux que s'il s'en remettait au hasard. La démonstration de (2) est plus une affaire de circonstances particulières : les sujets nient être conscients de tels événements, et leurs dénégations verbales sont confirmées par les données neurologiques qui établissent l'existence de lésions cérébrales d'un côté, et par la cohérence de ces dénégations d'autre part. C'est pourquoi nous les croyons ² !

Ce n'est pas un point trivial. Notez que ce qu'il y a de frappant au sujet de la vision aveugle s'évaporerait immédiatement si nous concluions que les sujets victimes de vision aveugle sont des simulateurs – et qu'ils prétendent seulement être conscients. Ou, plus près de la situation courante, comparez la façon dont nous acceptons les dénégations des sujets victimes de vision aveugle au scepticisme avec lequel nous accueillons ces mêmes dénégations quand elles viennent de personnes chez qui on a diagnostiqué « une cécité hystérique ». Il arrive que des gens dont les yeux et les cerveaux fonctionnent apparemment de façon normale (pour autant que les physiologues puissent le déterminer) se plaignent d'avoir été rendus aveugles ; ils confirment cette plainte en agissant « exactement comme des aveugles ». On peut souvent trouver une raison très plausible pour laquelle ce genre de sujets se motivent à « devenir » aveugles : ils veulent se punir eux-mêmes ou punir quelqu'un qui doit à présent s'occuper d'eux et se repentir de ne pas l'avoir fait, ou parce qu'ils veulent rejeter un souvenir visuel particulièrement terrible, ou comme une sorte de réponse panique à une autre maladie ou débilité. C'est donc

de la cécité « psychosomatique », si c'est vraiment de la cécité. Sont-ils réellement aveugles ? Il se peut qu'ils le soient. Après tout, pourrait-on dire, si la douleur psychosomatique est une vraie douleur, et si la nausée psychosomatique est assez vraie pour faire vomir quelqu'un, pourquoi la cécité psychosomatique ne pourrait-elle pas être une vraie cécité ?

Les aveugles par hystérie *prétendent* être aveugles, mais, comme les sujets victimes de vision aveugle, ils révèlent de façon tout à fait manifeste qu'ils *reçoivent de l'information visuelle*. Par exemple, il est significatif que les gens aveugles par hystérie réussissent *moins bien* qu'en se fiant au hasard si on leur demande de deviner divers traits visibles ! Cela montre de façon certaine qu'ils utilisent d'une façon ou d'une autre de l'information visuelle pour guider leur comportement de façon à faire un maximum d'« erreurs ». Les aveugles par hystérie ont le chic pour trouver des chaises sur lesquelles se cogner. Et pourtant, à la différence des purs simulateurs, quand les aveugles par hystérie disent qu'ils n'ont pas d'expériences visuelles, ils sont sincères — ils croient réellement ce qu'ils disent. Et nous, ne devrions-nous pas le croire aussi ? Comment devrions-nous traiter les textes de ces deux groupes différents de sujets quand nous extrapolons ce qu'il y a dans leurs mondes hétérophénoménologiques ?

Voilà un endroit où l'attitude extrêmement prudente de l'hétérophénoménologie est bénéfique. Les sujets victimes de vision aveugle et les aveugles par hystérie sont tous deux apparemment sincères dans leurs aveux quand ils disent ne pas avoir conscience que quoi que ce soit se produise dans leur champ visuel. Leurs mondes hétérophénoménologiques sont donc identiques — tout au moins relativement à ce champ visuel présumé aveugle. Et pourtant, il existe une différence. Nous avons une moins bonne connaissance des déterminants neuroanatomiques de la cécité hystérique que de la vision aveugle, et pourtant nous éprouvons, intuitivement, plus de scepticisme à leur égard quand ils nous font ces dénégations³. Qu'est-ce qui nous fait soupçonner que les aveugles par hystérie ne sont pas *réellement* aveugles et qu'ils sont d'une façon ou d'une autre ou à un degré quelconque conscients de leur univers visuel ? Ce qui nous fait nous poser des questions, ce sont les circonstances bizarrement propices de leur cécité, mais, au-delà de ces faits, il y a une raison simple : nous doutons de leurs affirmations quant à leur cécité parce que, *sans qu'on les y force*, les aveugles par hystérie *utilisent* quel-

quefois l'information que leurs yeux leur fournissent d'une façon différente de la manière dont les sujets victimes de vision aveugle utilisent cette information.

Il existe un facteur dans la situation expérimentale standard de la vision aveugle qui s'accorde si parfaitement avec nos hypothèses usuelles que presque personne ne se soucie de le discuter (voir cependant Marcel, 1988 ; van Gulick, 1989 ; Carruthers, 1989) : les sujets victimes de vision aveugle doivent être aiguillés ou recevoir des indications pour parvenir à faire leurs « conjectures » mieux que s'ils s'en remettaient au hasard. Ainsi, l'expérimentateur peut avoir dit, dans les instructions initiales : « Chaque fois que vous entendez un ton, faites une conjecture » ou « Chaque fois que vous sentez que je tape dans votre main, faites une réponse ». Sans ce genre d'indices, le sujet ne parvient pas à répondre ⁴.

Nous pouvons vérifier notre diagnostic de la différence en imaginant une variante. Supposez que nous ayons rencontré un sujet prétendument victime de vision aveugle à qui il ne faudrait pas donner d'indices : il « devine » spontanément (bien mieux qu'au hasard, mais pas parfaitement) qu'il y a quelque chose de présent dans le champ visuel supposé aveugle à chaque fois que cela se produit effectivement. Nous le laissons s'asseoir dans le laboratoire, et nous faisons les tests usuels pour repérer le scotome supposé ; il nous dit chaque fois quand la lumière qui avance disparaît dans son champ visuel aveugle, comme tous les autres sujets victimes de vision aveugle. Mais en même temps il émet spontanément des remarques du genre : « Ce n'est qu'une conjecture, mais ne venez-vous pas d'allumer une lumière dans mon scotome ? » Mais seulement après que nous venions de faire précisément cela. Cela paraît suspect, à tout le moins, et nous pouvons dire pourquoi.

En général, quand les sujets se conforment aux instructions dans une expérience, on considère que cela prouve indiscutablement qu'ils ont été capables de se conformer aux instructions parce qu'ils ont *consciemment fait l'expérience* des stimuli pertinents. C'est pourquoi l'instruction préparatoire suivante serait un non-sens :

Chaque fois que vous êtes *conscient* du fait que la lumière s'allume, pressez le bouton sur la gauche ; chaque fois que la lumière s'allume mais que vous n'êtes pas conscient du fait qu'elle s'allume, pressez le bouton sur la droite.

Comment un sujet pourrait-il se conformer à ce genre d'instruction ? Ce serait demander au sujet de faire quelque chose d'impossible : conditionner son comportement sur des occurrences qui lui seraient inaccessibles. Ce serait comme de dire « Levez la main chaque fois que quelqu'un vous fait un clin d'œil sans que vous vous en aperceviez ! »

Un expérimentateur ne ressentirait pas le besoin d'insérer l'adverbe « consciemment », comme dans :

Chaque fois que vous entendez consciemment un ton, faites une conjecture.

puisque l'hypothèse de base est que l'on ne peut pas conditionner ses propres politiques sur les expériences conscientes, même si cela arrive. Pour adopter la politique :

Chaque fois que x se produit, fais y

vous devez être conscient du fait que x se produit.

C'est notre hypothèse de base, mais cet édifice évident est fissuré. N'avons-nous pas entendu dire que nombre de nos comportements sont gouvernés par des conditions que nous ne détectons qu'inconsciemment ? Considérez les politiques qui règlent la température de nos corps, l'ajustement de notre métabolisme, le stockage et la dépense d'énergie, l'ajustement de notre système immunitaire ; considérez des politiques telles que cligner de l'œil quand des choses approchent ou entrent dans l'œil, et même des comportements publics tels que marcher (sans tomber) et baisser la tête quand des choses nous tombent soudain dessus. Tout ce « comportement » est contrôlé sans l'aide de la conscience, comme Descartes l'a bien vu.

Il semble, par conséquent, qu'il y ait deux sortes de politiques comportementales : celles qui sont contrôlées par la pensée consciente et celles qui sont contrôlées par des processus « aveugles, mécaniques » — tout comme les processus qui contrôlent un ascenseur automate. Un ascenseur automate n'est certainement pas conscient, et ne détecte rien consciemment ; il n'a donc pas de politiques conscientes. Mais on peut dire qu'il adhère à des politiques qui reposent sur divers états du monde qu'il détecte, et qu'il ajuste les politiques auxquelles il adhère sur la base d'autres états de choses qu'il détecte, et ainsi de suite. Il peut avoir des

politiques, des métapolitiques et des métamétapolitiques, qui reposent toutes sur diverses combinaisons d'états de choses détectés – toutes sans qu'il y ait le moindre état conscient. Quoi que puisse faire un ascenseur quand il détecte des états du monde ou suit une politique, un cerveau et un corps humains peuvent sûrement le faire aussi. Il peut suivre des politiques élaborées inconscientes comme celle que suit l'ascenseur.

Quelle est donc la différence entre suivre une politique inconsciente et suivre une politique consciente ? Quand nous considérons les politiques que nos corps suivent inconsciemment, grâce aux détecteurs « aveugles, mécaniques », il est tentant de dire que *puisque* ce sont des politiques inconscientes, ce ne sont pas tant *nos* politiques que celles de *nos corps*. *Nos* politiques sont (par définition, peut-on dire) nos politiques conscientes ; celles que nous formulons consciemment et délibérément, avec la possibilité de réfléchir (consciemment) sur le pour et le contre, et la possibilité de les ajuster ou de les amender au fur et à mesure que la situation évolue.

Il semble donc que quand on adopte initialement une politique à la suite d'une discussion verbale, ou en réponse à des instructions verbales, ce soit *ipso facto* une politique consciente, qui doit reposer sur des événements faisant l'objet d'une expérience consciente (Marcel, 1988). Ce qui semble contradictoire, c'est l'idée que l'on pourrait en parler puis décider de suivre une politique *inconsciente*, reposant sur des événements détectés inconsciemment. Mais il y a là, comme on peut le voir, une boucle : le statut de cette politique *doit* changer. Avec assez de pratique, et avec une capacité d'oubli stratégiquement bien placée, nous pourrions partir d'une politique adoptée consciemment puis nous diriger graduellement vers l'état consistant à suivre une politique inconsciente en détectant les diverses conditions sans en être conscients. Cela pourrait se produire, mais seulement si le lien avec les considérations verbales de la politique en question était en quelque manière brisé.

On peut mieux voir à quoi pourrait ressembler la transition possible dans l'autre direction. Un sujet victime de vision aveugle ne pourrait-il pas *devenir* conscient d'expériences visuelles dans le scotome, par un renversement du processus qu'on vient d'imaginer ? Après tout, dans la vision aveugle le cerveau du sujet reçoit manifestement l'information visuelle utilisée dans le travail de conjecture et l'analyse. Peu de temps après que se produit

le stimulus, quelque chose arrive dans le cerveau du sujet qui marque le début de l'état *informé*. Un observateur extérieur (comme l'expérimentateur) peut s'arranger pour reconnaître ces débuts « de la main gauche » bien qu'il ne soit pas conscient de ceux-ci « directement ». Dans ce cas, le sujet ne devrait-il pas être capable d'« éliminer le petit homme » et de venir à reconnaître, tout comme le fait l'expérimentateur, les changements qui interviennent dans ses propres dispositions ? Au départ, cela pourrait requérir l'emploi d'une sorte d'équipement de contrôle de soi – le même équipement que celui sur lequel s'appuie l'expérimentateur – mais avec le sujet qui regarde ou entend cette fois les signaux de sortie ⁵.

En d'autres termes, ne pourrait-il pas être en principe possible de « fermer la boucle de feedback » et par conséquent d'entraîner le sujet à suivre une politique de conditionnement de son propre comportement sur des changements dont il n'a pas (« directement ») fait l'expérience ? Je soulève la possibilité d'un tel entraînement dans la vision aveugle comme si c'était seulement une expérience de pensée. Nous pourrions essayer d'entraîner un sujet victime de vision aveugle à reconnaître quand il lui faut « deviner ».

Les sujets qui ont une vision aveugle ne sont pas immuables dans leurs capacités et leurs dispositions : certains jours ils sont en meilleure forme que d'autres ; ils s'améliorent avec la pratique, en dépit du fait qu'habituellement l'expérimentateur ne leur fournit *pas* de feedback en leur disant si leurs performances sont bonnes (pour des exceptions, voir Zihl, 1981). Il y a plusieurs raisons à cela : la principale est que toute situation expérimentale de ce genre est biaisée par quantité d'indices que l'expérimentateur peut fournir sans s'en rendre compte ; c'est pourquoi on minimise scrupuleusement et on contrôle les interactions entre l'expérimentateur et son sujet. Néanmoins, les sujets complètent les indices ou les suggestions qu'ils reçoivent de l'expérimentateur, et s'habituent graduellement à la pratique – qui serait en d'autres circonstances peu efficace – de faire des centaines ou des milliers de conjectures sur des sujets dont ils sont convaincus qu'ils n'ont pas la moindre expérience. (Imaginez ce que vous ressentiriez si l'on vous demandait de vous asseoir avec le Bottin du téléphone et de deviner quelle marque d'automobile possède chaque personne que vous trouvez sur la liste, sans qu'on vous dise jamais si vous avez deviné correctement. Cette activité ne paraîtrait pas motivée pendant très longtemps, à moins que vous n'ayez une assurance

incroyable sur la qualité de vos performances, et sur la raison pour laquelle un tel tour de force en vaille la peine.)

Supposez à présent que nous commencions à demander au sujet d'agir sans indices – « de deviner quand il faut deviner », de deviner « quand vous avez l'inspiration » – et à nouveau supposons que l'expérimentateur fournisse un feedback instantané. Il y a deux résultats possibles :

1. Le sujet commence au hasard et en reste là. En dépit du fait que le sujet soit informé (selon des modalités mesurables) du début du stimulus, il ne semble pas y avoir de façon dont il puisse découvrir quand cette information s'est produite, quelle que soit la quantité de « biofeedback » que nous lui accordions.

2. Le sujet devient finalement capable de travailler sans indices de la part de l'expérimentateur (ou du biofeedback temporaire), réussissant mieux que s'il se fiait au hasard.

Quel résultat obtiendrions-nous ? C'est bien entendu une question empirique, et je ne me risquerai même pas à chercher à deviner si un résultat du second type serait plausible. Peut-être que dans chaque cas particulier le sujet serait incapable de « deviner » correctement le moment où il lui faut deviner. Mais il faut noter que si un résultat de type 2 devait se produire, on pourrait alors très raisonnablement demander au sujet d'adopter des politiques qui requièrent qu'il relie son comportement aux stimuli dont l'occurrence serait seulement devinée par lui. Qu'il soit ou non conscient de ces stimuli, si sa fiabilité à « deviner » était élevée, il pourrait traiter ces stimuli sur le même plan que n'importe quelle expérience consciente. Il pourrait penser à et décider des politiques qui se relieraient à leur occurrence aussi facilement qu'à l'occurrence d'événements dont il fait consciemment l'expérience.

Mais cela le rendrait-il en quoi que ce soit *conscient* de ces stimuli ? Quelles sont vos intuitions ? Quand j'ai demandé aux gens ce qu'ils seraient enclins à dire dans ce genre de cas, j'ai obtenu des réponses mitigées. La psychologie populaire ne fournit pas de verdict clair. Mais un sujet victime de vision aveugle a parlé en son nom dans une circonstance particulière. D.B., l'un des sujets étudiés par Weizkranz, a une hémianopsie droite, et manifeste la capacité classique des sujets victimes de vision aveugle de deviner mieux qu'au hasard quand on lui donne des indices. Par exemple, si une lumière est lentement déplacée le long de son scotome horizontalement ou verticalement, et qu'il est poussé à deviner

« vertical ou horizontal », il réussit parfaitement bien, tout en niant avoir la moindre conscience du mouvement en question. Cependant, si la lumière est déplacée plus rapidement, elle devient son propre indice : D.B. peut se porter volontaire sans susciter un compte rendu très précis du mouvement, et même mimer le mouvement avec un geste de la main, dès qu'il survient (Weizkranz, 1988, 1989). Et quand on lui demande, D.B. insiste sur le fait que *bien sûr* il fait consciemment l'expérience du mouvement – sinon, comment serait-il capable de le rapporter ? (D'autres sujets qui ont une vision aveugle rapportent aussi une expérience consciente de stimuli qui se meuvent rapidement.) Nous pouvons vouloir réserver notre jugement, mais cette réponse ne devrait pas nous surprendre, si l'analyse de Rosenthal du concept ordinaire de conscience est correcte. D.B. n'est pas simplement informé du mouvement de la lumière ; il *comprend* qu'il a été informé ; dans les termes de Rosenthal, il a une pensée du second-ordre, qui a pour contenu qu'il vient d'avoir la pensée du premier-ordre.

Otto, notre critique, rétorque :

Mais c'est encore un tour de passe-passe ! Nous savons depuis toujours que les sujets victimes de vision aveugle étaient conscients de deviner. Tout ce que cela montre, c'est qu'un tel sujet pourrait acquérir une capacité de deviner le moment où il lui faut deviner (et il serait bien entendu conscient de deviner dans ces cas). Le fait de reconnaître les conjectures de quelqu'un sur ces sujets pourrait difficilement par soi-même suffire pour devenir directement conscient de l'événement que l'on est en train de deviner.

Cela suggère qu'on a encore besoin de quelque chose pour qu'il y ait conscience visuelle. Que pourrait-on ajouter ? D'abord le lien entre une conjecture et l'état sur lequel il *porte*, bien que fiable, semble très mince et éphémère. Ne pourrait-on l'assurer et le renforcer ? Que serait le résultat si les liens qui font qu'une conjecture est *à propos de* son objet (*aboutness*) étaient multipliés ?

Cache-tampon : un exercice de prise de conscience

Le terme philosophique standard désignant la propriété d'être à propos de quelque chose est celui d'*intentionnalité*, et selon Elizabeth Anscombe (1965), c'est un terme qui « provient méta-

phoriquement » du latin *intendere arcum*, ce qui veut dire *viser* (quelque chose) *avec un arc et une flèche*. Cette image de visée ou de direction est au centre de la plupart des discussions philosophiques sur l'intentionnalité, mais en général les philosophes ont échangé le processus complexe de viser avec une flèche réelle pour une flèche purement « logique », une relation fondamentale ou primitive, rendue d'autant plus mystérieuse qu'elle est simple. Comment quelque chose qui est dans votre tête pourrait-il viser avec cette flèche abstraite une chose du monde ⁶ ? Il peut en dernière instance être correct de traiter la propriété d'être-à-propos-de comme une relation abstraite, logique. Mais, au départ, cela détourne l'attention des processus qui sont effectivement impliqués dans le maintien d'un esprit en contact avec les choses du monde de façon à ce que celles-ci soient effectivement visées par la pensée : les processus d'attention, de conservation de contact, de suivi à la trace et de recherche des objets (Selfridge, à paraître). Le processus effectif de viser quelque chose, de « la garder dans son viseur » implique que l'on fasse une série d'ajustements et de compensations dans le temps, sous un « contrôle feedback ». C'est pourquoi la présence de nuées de distracteurs (comme le ruban métallique antiradar qui confond les systèmes de détections antimissiles) peuvent rendre la visée impossible. Bloquer une cible suffisamment longtemps pour l'identifier est un exploit qui requiert plus qu'une transaction d'information unique et momentanée. Le meilleur moyen de rester en contact avec quelque chose est, littéralement, de garder le *contact* avec elle – de la saisir et de ne pas la laisser partir, en sorte que l'on puisse l'examiner au fond de son cœur, en prenant son temps. Le second meilleur moyen est de rester en contact avec elle figurativement, en la suivant à la trace des yeux (et du reste de son corps), en ne la perdant jamais de vue. C'est quelque chose qui peut être accompli par la perception, bien entendu, mais pas simplement par la perception passive ; cela peut requérir de l'effort, de la planification, et en tout cas une *activité continue* de maintien du contact.

Quand j'étais enfant, j'adorais jouer à cache-tampon. Un tampon ordinaire est montré à tous les participants, et tous sauf un quittent la pièce, pendant qu'on « cache » le tampon. Les règles pour celui qui le cache sont claires : le tampon doit être caché *à la vue de tous*. Il ne peut pas être placé derrière ou sous quoi que ce soit, ou trop haut pour que les enfants puissent le voir.

Dans le salon moyen, il existe des douzaines de lieux où vous pouvez placer un tampon et où il tendra à se mélanger aux objets qui l'entourent comme un animal bien camouflé. Une fois qu'il est caché, les autres enfants reviennent dans la pièce et se mettent à chasser le tampon. Dès que vous voyez le tampon, vous vous asseyez sagement, en essayant de ne pas trahir sa localisation. On peut en général s'attendre à ce que les autres enfants qui n'ont pas encore trouvé le tampon regardent *directement* dans la direction du tampon sans en fait le *voir*. Dans ces moments délicieux, tous les autres peuvent voir que le tampon est directement en face du nez de Marinette – par exemple – bien éclairé et sous un bon angle dans son champ visuel. (À ces moments-là, ma mère aimait dire : « Si c'était un ours, il te mordrait ! ») En entendant les gloussements et les cris de surprise des autres enfants, Marinette pourrait elle-même réaliser qu'elle doit être droit devant – et pourtant qu'elle ne le voit toujours pas.

Nous pourrions le formuler ainsi : même si un état représentationnel quelconque dans le cerveau de Marinette « inclut » en quelque façon le tampon, aucun état perceptuel de Marinette n'est encore *à propos* du tampon. Nous pouvons admettre que l'un de ses états conscients porte sur le tampon : son « image de recherche ». Elle peut ardemment se concentrer sur la découverte du tampon, le tampon qu'elle a été autorisée à examiner il y a une minute ou deux. Mais aucune relation forte d'intentionnalité ou d'être-à-propos-de n'existe entre ses états perceptifs et le tampon, bien qu'elle puisse avoir l'information dans un état quelconque de son système visuel, qui donnerait à quelqu'un d'autre (un observateur extérieur par exemple, qui étudierait les états de son cortex visuel) la possibilité d'identifier le tampon. Marinette doit « faire un réglage » sur le tampon, elle doit le séparer comme « figure » et fond » et l'identifier. Lorsque c'est fait, Marinette voit réellement le tampon. Le tampon sera finalement « dans son expérience consciente » – et à présent qu'elle en est consciente, elle sera enfin en mesure de lever sa main d'un geste de triomphe – ou de s'asseoir calmement avec les autres enfants qui ont déjà repéré le tampon ⁷.

Ce genre de liens intentionnels guidés par des feedback, où les erreurs sont corrigées, les gains ajustés, est requis pour que puisse exister la sorte de familiarité avec les objets qui soit digne de ce nom, qui puisse servir de repère pour conduire une certaine politique, par exemple. À partir du moment où j'ai quelque chose

en ce sens fort, je « peux faire quelque chose à son sujet » ou je peux faire quelque chose *parce que* je l'ai vu ou *dès que* je le vois. Les tampons individuels, une fois identifiés, sont en général assez faciles à suivre à la trace (à moins bien sûr que vous vous trouviez dans une pièce pleine de tampons pendant un tremblement de terre). Dans des circonstances normales, par conséquent, le statut qu'a acquis le tampon dans le système de contrôle de Marinette ne dure pas simplement pendant un moment fugitif : le tampon restera localisé par Marinette pendant la recherche, ou pendant le temps nécessaire pour réaffirmer son identité, pour le vérifier encore (et encore – s'il y a des raisons d'en douter). Les choses dont nous sommes en définitive conscients sont les choses que nous observons franchement et sans précipitation, en rassemblant et en assimilant les fruits de nombreuses saccades, construisant une familiarité dans le temps tout en gardant l'objet localisé dans notre espace personnel. Si l'objet se balade comme un papillon, nous nous mettrons en fait à l'immobiliser, « afin de pouvoir le regarder », et s'il est bien camouflé dans ses alentours, il nous faudra avancer d'un pas – littéralement, si nous ne devons pas le toucher – pour l'obtenir sur un arrière-plan contrasté.

Si nous n'y parvenons pas, cela peut nous empêcher de voir l'objet, au sens usuel de ce terme ⁸.

Les observateurs d'oiseaux gardent souvent une liste de toutes les espèces qu'ils ont vues durant leur existence. Supposez que nous soyons des observateurs d'oiseaux et que nous entendions tous deux un oiseau chanter au-dessus de nos têtes dans les arbres. Je lève la tête et je dis : « Je le vois. Pas toi ? » Vous regardez exactement dans la même direction que moi, et pourtant vous dites, de bonne foi : « Non, je ne le vois pas. » Je me mets à écrire le nom de cet oiseau sur ma liste ; vous ne le faites pas, en dépit du fait que vous puissiez être moralement certain que son image doit avoir passé de façon répétée devant vos fovéas.

Que diriez-vous ? Le tampon n'était-il pas en quelque sorte « présent » dans la conscience de Marinette avant qu'elle l'ait détecté ? L'oiseau n'était-il pas présent dans l'« arrière-plan » de votre conscience, ou pas présent du tout ? Placer quelque chose sur le devant de votre conscience c'est le placer dans une position où on peut rapporter à son sujet, mais que faut-il pour que quelque chose se place dans l'arrière-plan de votre expérience consciente (et non pas simplement dans l'arrière-plan de votre

environnement) ? Le tampon et l'oiseau étaient sans nul doute présents dans l'environnement visible – ce n'est pas le problème. On peut présumer qu'il ne suffit pas que la lumière réfléchie par l'objet entre seulement dans vos yeux, mais quel doit être l'effet supplémentaire que doit avoir la lumière réfléchie – quel enregistrement supplémentaire doit en avoir votre cerveau – pour que l'objet passe du rang de ce qui fait l'objet d'une réponse inconsciente au rang d'objet figurant dans l'arrière-plan de votre expérience consciente ?

Comment répondre à ces déclarations faites du « point de vue de la première personne » ? En ignorant le point de vue de la première personne, et en examinant ce que l'on peut apprendre du point de vue de la troisième personne. Aux chapitres 7 à 9, nous avons exploré un modèle de production d'actes de langage qui dépendait d'un processus de Pandémonium dans lequel l'accouplement final des contenus avec les expressions était le terme d'une série de compétitions, de constructions, puis de démantèlements et de reconstructions de coalitions. Les contenus qui entraient dans la mêlée mais qui ne parvenaient pas à se perpétuer longtemps pouvaient envoyer une sorte d'effet à un coup « balistique » se propageant en onde à travers le système, mais ils étaient quasiment impossibles à *rapporter*. Quand un événement ne s'attarde pas, toute tentative pour le rapporter, si on s'y met, ou bien avortera ou bien vagabondera en dehors de tout contrôle, faute de quelque appui à partir duquel elle pourrait se corriger. Pour qu'un événement puisse être rapporté, il faut qu'il y ait une capacité à identifier et à réidentifier l'effet. Nous pouvons voir comment cette capacité à rapporter quelque chose se développe dans de nombreuses variétés d'entraînement, qui rappellent l'entraînement que nous imaginons donner à nos patients victimes de vision aveugle : les résultats d'un entraînement du palais chez les taste-vin, l'entraînement des oreilles chez les musiciens, et autres choses du même genre – ou l'expérience simple du pincement des cordes de guitare décrite au chapitre 2.

Considérons, par exemple, les instructions que l'on donne aux accordeurs de pianos débutants. On leur demande d'écouter les « mesures » quand ils frappent les touches qu'ils accordent selon une touche de référence. Au début, la plupart des novices sont incapables de discerner dans leur expérience auditive quoi que ce soit qui réponde à la description « mesure » – ce qu'ils entendent est quelque chose qu'ils décriraient comme une sorte de mauvaise

résonance sans structure ou sonnait faux. Mais finalement, si l'entraînement réussit, ils viennent à être capables d'isoler, dans leur expérience auditive, l'interférence « mesure », et de noter comment les trames de mesures changent en réponse à leurs tours sur le « marteau » accroché sur la « cheville ». Ils peuvent alors facilement accorder le piano en accordant les mesures. Ce qu'ils disent en général – et nous pouvons tous le confirmer avec des épisodes semblables dans notre propre expérience –, c'est qu'en vertu de leur entraînement *leur expérience consciente a changé*. Plus spécifiquement, elle a augmenté : ils sont à présent conscients de choses dont ils n'étaient pas auparavant conscients.

Or, en un sens, bien entendu, ils ont entendu les mesures depuis le début. C'est l'interférence, après tout, qui compose les sons faux dont ils étaient certainement conscients. Mais ils étaient auparavant incapables de détecter ces composants dans leur expérience. C'est pourquoi on peut dire que ces facteurs ont *contribué* à l'expérience, mais n'étaient pas eux-mêmes présents dans celle-ci. Le statut fonctionnel de ces contributions préalables à l'apprentissage était le même que celui des événements qui se produisent dans la vision aveugle : le sujet est incapable de rapporter les contributions particulières, ou de mettre en place une politique quand celles-ci débudent, mais les résultats de ces contributions peuvent toujours être rendus manifestes dans le comportement du sujet, par exemple dans sa capacité à répondre à des questions habilement posées. Ce que je suis en train de suggérer est que c'est à cela que revient le fait d'être dans l'arrière-plan de l'expérience. Mais il n'est pas hors de question, comme nous l'avons vu, qu'un lien plus fort du type de celui que nous avons décrit pour les accordeurs de pianos et pour les taste-vin puisse se construire chez le sujet victime de vision aveugle, en sorte qu'il puisse déclarer, et puisse aisément accepter, qu'il soit *devenu* conscient de stimuli – même sur le devant de sa conscience – qu'il ne pouvait au départ que conjecturer.

Pas si vite [dit Otto]. Voici une autre objection : tu imagines que le sujet victime de vision aveugle apprend à utiliser ses capacités de vision aveugle de nouvelles façons, et que peut-être cela lui donnera une *sorte* de conscience des événements qui se produisent dans son champ visuel aveugle, mais cela laisse toujours quelque chose de côté. La conscience ne serait pas la conscience *visuelle* ; ce ne serait pas comme *voir*. Les « qualités phénoménales » ou les *qualia* de la vision consciente seraient absentes, même si le sujet

qui a une vision aveugle pouvait faire toutes ces opérations fonctionnelles.

Peut-être, et peut-être pas. Que sont donc les « qualités phénoménales » ou les *qualia* (*qualia* est le terme latin pour « qualité » ; le singulier est *quale*) ? Elles semblent terriblement évidentes au départ – ce sont les façons dont les choses nous apparaissent, dont elles sont senties, sont éprouvés ou entendues par nous, mais elles ont une façon de changer de statut ou de s'évanouir quand nous les examinons. Dans le chapitre suivant, nous suivrons ces suspects à la trace jusqu'à quelques bosquets philosophiques, mais d'abord il nous faut considérer quelques propriétés qui ne sont *pas* des qualités phénoménales, mais qui pourraient être confondues avec elles.

La vision prothésique : à part l'information, qu'est-ce qui manque ?

Le sujet de Weizkranz, D.B., *voit-il* le mouvement ? Bon, il ne l'*entend*, ni ne le *sent*, sûrement pas. Mais est-ce de la vision ? A-t-il les « qualités phénoménales » de la vision ? Weizkranz dit :

« Au fur et à mesure que la “ saillance ” du stimulus augmente, le patient peut dire de façon insistante qu'il ne “ voit ” toujours pas, mais il a désormais une sorte de “ sentiment ” qu'il y a quelque chose. Dans certains cas, si la saillance est augmentée un peu plus, on peut atteindre un point où le sujet dit qu'il “ voit ” mais l'expérience n'est pas véridique. Par exemple, D.B. “ voit ” en réponse à un stimulus qui se meut vigoureusement, mais il ne le voit pas comme un objet cohérent en mouvement, mais rapporte plutôt des trames complexes d'“ ondes ”. D'autres sujets rapportent des “ ombres sombres ” qui émergent au fur et à mesure que la luminosité et le contraste augmentent à un haut degré » (1988, p. 189).

L'objet qui se meut vigoureusement n'est pas perçu par D.B. comme ayant une couleur ou une forme, mais qu'est-ce que cela prouve ? Comme nous nous en sommes convaincus au chapitre 1, dans l'expérience de la carte à jouer tenue dans la vision périphérique, nous pouvons certainement voir la carte sans être capable d'identifier soit sa couleur, soit ses formes. C'est de la vision normale, pas de la vision aveugle, et c'est pourquoi nous devrions pour ces raisons refuser de dénier au sujet une expérience visuelle.



Figure 10.3 : Un sujet aveugle avec un système portable électrique à 16 lignes. La caméra télé est attachée à un objectif, monté sur une paire de cadres de lunettes. Un petit amas de fils est relié à un circuit électrique chargé de conduire les stimuli (tenu dans sa main droite). La matrice de 256 électrodes d'argent est tenue dans sa main gauche.

De *Brain Mechanisms in Sensory Substitution*, Paul Bach-y-Rita, © 1972, Academic Press Inc.

On peut poser de façon plus frappante la question de savoir si cette façon anormale d'obtenir de l'information sur les choses visibles serait une variété de *vision*, en envisageant une forme de « vision » plus radicalement différente encore. On a construit des prothèses pour fournir de la « vision » aux aveugles, et certains de ces dispositifs soulèvent les mêmes problèmes. Il y a près de vingt ans, Paul Bach-y-Rita (1972) a mis en place plusieurs dispositifs qui consistaient en de petites caméras vidéo de très faible résolution qu'on pouvait monter sur des cadres de lunettes. Le signal de faible résolution provenant de ces caméras, une série de pixels « noirs et blancs » de 16 sur 16 ou de 20 sur 20, était

répandu sur le dos ou sur le ventre du sujet sur une grille de picotements à vibration électrique ou mécanique – les « tacteurs ».

Après seulement quelques heures d'entraînement, les sujets aveugles munis de ce dispositif pouvaient apprendre à interpréter les structures de picotements qu'ils ressentaient sur leur peau, à peu près comme vous pouvez interpréter des lettres tracées par quelqu'un sur votre peau avec son doigt. La résolution est faible, mais malgré cela les sujets pouvaient apprendre à lire des signes, et à identifier des objets et même des visages, comme nous pouvons nous en apercevoir en *regardant* cette photo du signal telle qu'elle apparaît sur le moniteur d'un oscilloscope.

Le résultat a certainement produit avec une prothèse une expérience de perception consciente, mais puisque l'entrée était

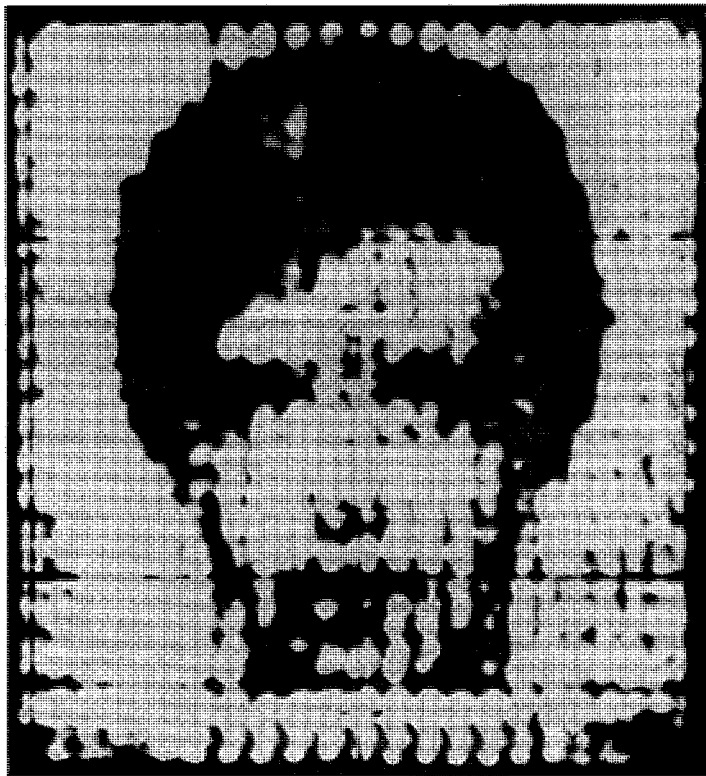


Figure 10.4 : Apparition d'une représentation à résolution 400 d'une figure de femme sur le moniteur d'un oscilloscope. Les sujets peuvent identifier correctement des trames de stimuli de ce niveau de complexité.

répandue sur le dos ou sur le ventre des sujets au lieu de leurs rétines, était-ce de la *vision* ? Avait-elle les « qualités phénoménales » de la vision, ou était-ce seulement une sensation tactile ?

Rappelez-vous l'une de nos expériences au chapitre 2. Votre point de vue tactile peut très aisément s'étendre jusqu'au bout d'un stylo, vous permettant de ressentir des textures avec le bout, tout en oubliant les vibrations du stylo contre vos doigts. Nous ne serions donc pas surpris si nous apprenions que les sujets de Bach-y-Rita éprouvent un effet similaire bien que plus extrême. Après une brève période d'apprentissage, leur conscience des picotements sur leur peau disparaissait : le bloc de pixels devenait transparent, pour ainsi dire, et le point de vue du sujet se déplaçait sur le point de vue de la caméra, montée à côté de leurs visages. Une confirmation frappante de ce changement de point de vue était le comportement d'un sujet d'expérience dont la caméra avait un objectif zoom avec un bouton de contrôle (p. 98-99). La série des picotements était sur son dos, et la caméra était montée sur le côté de sa tête. Quand l'expérimentateur touchait le bouton de zoom sans prévenir, provoquant l'expansion ou le surgissement de l'image sur le *dos* du sujet, le sujet reculait instinctivement, *levant le bras pour protéger sa tête*. Une autre démonstration frappante de la transparence des picotements est le fait que les sujets qui avaient été entraînés avec la série de picotements sur leur dos pouvaient s'adapter immédiatement quand les picotements étaient déplacés vers leur ventre (p. 33). Et pourtant, comme le note Bach-y-Rita, ils répondaient toujours à une démangeaison sur le dos comme à quelque chose à gratter – ils ne se plaignaient pas de la « voir » – et ils étaient parfaitement capables de prêter attention aux picotements, en tant que picotements, quand on le leur demandait.

Ces observations sont fascinantes, mais elles ne sont pas concluantes. On pourrait soutenir qu'à partir du moment où l'usage des entrées d'un certain dispositif était devenu une seconde nature, les sujets étaient réellement en train de voir, ou, inversement, que seuls certains des traits fonctionnels les plus centraux du « voir » avaient été reproduits prothésiquement. Qu'étaient devenues les autres « qualités phénoménales » de la vision ? Bach-y-Rita rapporte les résultats de l'expérience suivante : montrer à deux sujets aveugles entraînés, des étudiants mâles de première année d'université, pour la première fois de leur vie, des pho-

tographies de femmes nues tirées de *Playboy*. Ils furent déçus – bien que tous deux purent décrire une bonne partie du contenu des photos, l'expérience n'avait pas de contenu d'affect : aucune sensation agréable ne leur vint. Cela troubla beaucoup ces deux jeunes gens, qui étaient conscients du fait que des photos similaires avaient un contenu d'affect pour leurs camarades dotés d'une vision normale (p. 145).

Les dispositifs prothésiques de Bach-y-Rita ne produisaient donc pas *tous* les effets de la vision normale. Une partie du déficit doit être due à la différence énorme dans la vitesse du flux d'information. La vision normale nous informe sur les propriétés spatiales des choses de notre environnement à une grande vitesse et avec à peu près autant de détails que nous en avons envie. Il n'est pas surprenant qu'une information spatiale de faible résolution envoyée vers le cerveau à travers un interface sur la peau n'ait pas réussi à provoquer toutes les réactions qui sont provoquées chez les personnes dont la vision est normale quand leurs systèmes visuels sont envahis par cette entrée d'information⁹. Quelle quantité de plaisir devrions-nous attendre d'une personne douée de vision normale puisse tirer du fait qu'elle *regarde* des traductions semblables de faible résolution – il n'y a qu'à regarder la figure 10.4 – d'images de personnes agréables à regarder ?

On ne voit pas très bien ce que cela changerait si nous pouvions d'une façon ou d'une autre changer le « taux de bauds »¹⁰ de la vision prothésique pour qu'elle puisse correspondre à la vision normale. Il se pourrait qu'augmenter simplement la quantité et le taux de l'information, en fournissant en quelque façon des cartes à bits de plus haute résolution au cerveau, suffirait à produire l'agrément qui manque. Ou plutôt une partie de cet agrément. Les aveugles de naissance seraient terriblement désavantagés par rapport aux gens qui ont récemment perdu la vue, parce qu'ils n'ont construit aucune des associations spécifiquement visuelles qui jouent sans aucun doute un rôle important dans l'agrément que prennent les gens qui ont une vue normale dans leurs expériences. Il se pourrait aussi que le plaisir que nous prenons aux expériences visuelles soit le sous-produit d'anciennes traces fossilisées d'une économie antérieure de nos systèmes nerveux. C'est là une question que nous avons soulevée au chapitre 6 et sur laquelle nous reviendrons dans le chapitre suivant.

Les mêmes considérations s'appliquent à la vision aveugle et à toutes les améliorations que nous pourrions envisager des capacités des sujets qui ont une vision aveugle. Les discussions de la vision aveugle ont tendu à ignorer combien l'information que les sujets peuvent récupérer de leurs champs visuels aveugles est misérable. C'est une chose que d'être capable de deviner, quand on vous donne des indications, si on vous a présenté dans votre champ aveugle un cercle ou un carré. Ce serait une tout autre question que d'être capable de deviner dans le détail, quand on vous donne des indications, ce qui vient de se passer de l'autre côté de la fenêtre.

Nous pouvons utiliser ce que nous avons appris de la vision prothésique pour guider notre imagination quant à l'effet que cela ferait à un sujet victime de vision aveugle de recouvrer plus de *fonctions* de la vision. Essayons d'imaginer que nous rencontrons une personne corticalement aveugle qui, après un entraînement assidu, (1) a transformé sa capacité à deviner au moment opportun en une seconde nature, (2) peut jouer à cache-tampon avec les meilleurs, et (3) a réussi d'une façon ou d'une autre à augmenter le détail de ses conjectures de façon importante. Nous la rencontrons en train de lire le journal et de sourire en lisant les bandes dessinées, et nous lui demandons de s'expliquer. Il existe trois scénarios possibles, du moins plausible au plus plausible :

(1) « Je me suis contenté de deviner, bien entendu ! » Je ne peux pas voir la moindre chose, vous savez, mais j'ai appris à deviner quand il le fallait, et en ce moment, par exemple, je suppose que vous êtes en train de faire un geste grossier, et que vous grimacez en signe d'incrédulité totale ! »

(2) « Eh bien, ce qui a commencé comme une simple conjecture a finalement perdu ce statut, au fur et à mesure que je me suis mis en confiance. La conjecture est devenue *pressentiment*, pour ainsi dire. Je *savais* soudain que quelque chose se passait dans mon champ visuel. Je pouvais alors exprimer ma connaissance, et agir sur elle. De plus j'avais alors une métaconnaissance que j'étais en fait capable de ce genre de pressentiments, et de ce que j'étais capable d'utiliser cette métaconnaissance pour planifier mes actions et instaurer des politiques. Ce qui commençait par être des conjectures conscientes devenait des sentiments conscients, qui sont à pré-

sent si rapides et furieux que je ne peux même pas les séparer. Mais je ne parviens toujours pas à voir quoi que ce soit ! Pas comme je voyais avant ! Ce n'est pas du tout de la vision ! »

(3) « Eh bien, en fait, c'est *très proche* de la vision. J'agis à présent sans effort dans le monde sur la base de l'information recueillie par mes yeux sur mon environnement. Ou je peux avoir une conscience réflexive de ce que j'obtiens de mes yeux si j'en ai envie. Sans la moindre hésitation je réagis aux couleurs des choses, à leur forme, à leur localisation, et j'ai perdu tout sens de l'effort que j'ai dépensé pour développer ces talents et en faire une seconde nature. »

Et pourtant, nous pouvons toujours imaginer que notre sujet dit que quelque chose manque encore :

« *Les qualia*. Mes états de perception ont bien des qualia, bien sûr, parce que ce sont des états conscients, mais avant que j'aie perdu la vue, ces états avaient des qualia *visuels*, alors qu'à présent ils n'en ont pas, en dépit de tout l'entraînement que j'ai reçu. »

Il peut vous sembler évident que ce soit cohérent, que ce soit juste ce à quoi vous puissiez vous attendre de la part du sujet. Si c'est le cas, le reste du chapitre est pour vous, et c'est un exercice destiné à ébranler cette conviction. Si vous doutez déjà que ce discours sur les qualia ait la moindre signification, vous avez probablement anticipé les ressorts de mon argument.

« Remplir » contre trouver

« Mais l'existence de ce sentiment d'étrangeté ne nous donne pas de raison de dire que tout objet que nous connaissons bien et qui ne nous paraît pas étrange nous procure un sentiment de familiarité.

Nous pensons que, pour ainsi dire, la place remplie à un certain moment par le sentiment d'étrangeté doit sûrement être occupée *en quelque façon*. »

Wittgenstein, 1953.

Au chapitre 1, nous avons vu que l'une des raisons de croire au dualisme était qu'il promettait de nous fournir « la matière dont les rêves sont faits » – les vaches violettes et autres produits de notre imagination. Et au chapitre 4, nous avons vu les confu-

sions qu'engendrait l'hypothèse naturelle mais erronée selon laquelle *après* que le cerveau est parvenu à une discrimination ou à un jugement, il *re-présente* le matériau sur lequel se fonde le jugement, pour le plaisir de l'auditoire du Théâtre Cartésien, en remplissant le tout de couleurs vives. L'idée de *remplissement*, qui est commune même chez les théoriciens subtils, est une fausse bonne idée, vestige du matérialisme cartésien. Ce qu'il y a d'amusant est que ceux qui utilisent ce terme en savent souvent plus, mais que, comme ils trouvent ce terme irrésistible, ils se couvrent en le mettant entre guillemets.

Par exemple, tout le monde décrit le cerveau comme « remplissant » la tache aveugle (dans tous les exemples, c'est moi qui mets les italiques) :

« Le phénomène neurologique bien connu du "*remplissement*" subjectif de la portion manquante d'une aire aveugle dans le champ visuel » (Libet, 1985b, p. 567).

« Vous pouvez localiser votre propre tache aveugle, et aussi démontrer comment une trame est "*remplie*" ou "*complétée*" dans la tache aveugle » (Hundert, 1987, p. 427).

Il y a aussi le « remplissement » auditif. Quand nous entendons un discours parlé, des lacunes dans le signal acoustique peuvent être « remplies » – par exemple dans le phénomène de restauration du phonème » (Warren, 1970). Ray Jackendof le formule ainsi :

« Envisageons, par exemple, la perception du discours quand les entrées sont défectueuses ou contiennent du bruit – par exemple en présence d'un avion en vol ou pendant une mauvaise connexion téléphonique [...] Ce que l'on construit [...] n'est pas simplement une signification voulue mais aussi une structure phonologique : on "entend" plus que ce que le signal ne véhicule en fait... En d'autres termes, l'information phonétique est "*remplie*" à partir de structures de niveau supérieur tout autant qu'à partir du signal acoustique ; et bien qu'il y ait une différence dans la façon dont il est dérivé, il n'y a pas de différence qualitative dans la structure complétée elle-même » (Jackendof, 1989, p. 99).

Et quand nous lisons un texte, quelque chose de similaire se produit (sur le plan visuel). Bernard Baars le formule ainsi :

« Nous trouvons des phénomènes similaires dans l'effet bien connu du "correcteur d'épreuves", où la recherche générale de coquilles

dans les pages d'épreuves est difficile à détecter parce que l'esprit "remplit" le texte avec l'information correcte » (Baars, 1988, p. 17).

Howard Margolis ajoute un commentaire incontestable sur ce processus de « remplissement » :

« Les détails qui font l'objet d'un "remplissement" sont en général corrects » (Margolis, 1987, p. 41).

La reconnaissance tacite du fait qu'il y a quelque chose qui cloche quand on parle de « remplissement » se manifeste bien dans cette description de la tache aveugle qu'on trouve chez le philosophe C.L. Hardin, dans son livre *Colors for Philosophers* :

« Elle couvre une aire de six degrés de diamètre, suffisamment pour contenir les images de dix lunes pleines placées côte à côte, et pourtant il n'y a pas de trou dans la région correspondante du champ visuel. C'est parce que l'œil-cerveau la *remplit* de tout ce qu'il voit dans les régions adjacentes. Si ceci est bleu, il le *remplit* de bleu ; si cette surface est de couleur écossaise, nous ne percevons aucune discontinuité dans l'écossais » (1988, p. 22).

Hardin ne parvient pas à dire que le cerveau remplit la surface avec de l'écossais, parce que cela suggère, sans aucun doute, une forme sophistiquée de « construction », comme le « raccommodage invisible » ingénieux qu'on peut vous faire sur votre veste à chevrons, et qui peut vous coûter fort cher : toutes les lignes s'alignent, et toutes les nuances de couleur s'insèrent dans la frontière entre l'ancien et le nouveau morceau de tissu. Il semble que remplir de couleur bleue soit une chose – il suffit d'un ou deux petits coups de pinceau cérébral avec la bonne couleur ; mais remplir en écossais est autre chose, et c'est plus que Hardin ne peut s'autoriser.

Mais comme la remarque de Hardin nous le rappelle, nous oublions tout autant nos taches aveugles quand nous sommes en présence d'un champ écossais que quand nous sommes en présence d'une surface de couleur uniforme, en sorte que le processus qui produit cet oubli peut aussi facilement être créé par le cerveau dans chaque cas. « Nous ne sommes conscients d'aucune discontinuité », comme il dit. Mais si le cerveau n'a pas à remplir le trou avec de l'écossais, pourquoi s'occuperait-il de le remplir de bleu ?

À aucun moment, semble-t-il, le « remplissement » n'est un remplissement au sens littéral – au sens où il faudrait quelque

chose comme des pinceaux (c'était la morale de l'histoire du CADAVEUGLE de Série II au chapitre 9). Je suppose que personne ne pense que le « remplissement » ne porte sur une activité du cerveau qui se fatiguerait à couvrir une surface spatiale avec un *pigment*. Nous savons que l'image réelle, inversée, qui est sur la rétine est la dernière étape de la vision dans laquelle il y ait quoi que ce soit de coloré comme quand une image sur un écran est normalement colorée. Puisqu'il n'y a pas d'œil de l'esprit au sens littéral, on n'a pas besoin de pigment dans le cerveau.

Voilà pour le pigment. Cela ne nous empêche pas d'être encore tentés de penser qu'il se passe quelque chose dans le cerveau qui soit *en quelque façon* analogue à la coloration d'une surface par un pigment — autrement nous ne parlerions pas de « remplissement ». C'est cet événement spécial, quel qu'il soit, qui se produit dans le « véhicule » spécial de l'expérience visuelle ou auditive, semble-t-il. Comme le dit Jackendof, au sujet de l'expérience auditive, on « entend » plus que ce que le signal ne transporte effectivement. Notez cependant qu'il met encore « entend » entre guillemets. Qu'est-ce qui pourrait bien être *présent* quand on « entend » des sons remplir des silences ou quand on « voit » des couleurs se répandre sur des espaces vides ? Il semble que quelque chose soit là dans ces cas, quelque chose que le cerveau ait à fournir (en « remplissant »). Quel nom pourrions-nous donner à ce quoi-que-ce-soit ? Appelons-le un *figment*. Nous sommes donc tentés de supposer qu'il y a quelque chose, fait de figment, qui se trouve là quand le cerveau « remplit » et qui n'est pas là quand il ne s'occupe pas de « remplir ». Aussi mal formulée, l'idée de figment n'aura guère d'attrait. (Du moins j'espère qu'elle n'en a pas.) Nous avons mieux : il n'existe pas de matière de ce genre. Le cerveau ne produit pas de figment ; le cerveau n'utilise pas de figment pour remplir les lacunes ; le figment n'est qu'un figment de mon imagination. Voilà ce qu'il en est du figment ! Mais alors que veut dire « remplir », qu'est-ce que cela *pourrait* bien vouloir dire, si cela ne veut pas dire remplir de figment ? S'il n'existe pas de véhicule tel que le figment, en quoi le « remplissement » se distingue-t-il du fait de ne pas se soucier de remplir quoi que ce soit ?

Au chapitre 9, nous avons vu comment un système CAD pouvait représenter des couleurs en associant un nombre de couleur à chaque pixel, ou à chaque région délimitée de l'objet dépeint, et nous avons vu comment le CAD aveugle de Série II pouvait recher-

cher, ou détecter, des couleurs en lisant un tel code. Ce processus rappelle le passe-temps pour les enfants, colorier avec les nombres, qui nous fournit un analogue des processus de représentation qui doivent se produire, ou qui pourraient se produire dans le cerveau. La figure 10.5 est une représentation qui contient des informations sur les formes, mais pas sur les couleurs. Comparez-la avec la figure 10.6, qui contient des informations sur les couleurs, sous la forme d'un code numérique. Si vous deviez prendre des crayons et suivre les directives de remplissage des couleurs, vous transformeriez la figure 10.6 en une autre sorte de représentation « remplie » – dans laquelle les régions seraient remplies de véritables couleurs, de vrai pigment.

On peut encore « remplir » de couleur d'une autre manière, pixel par pixel, au moyen d'une carte à bits codée, comme dans la figure 10.7.

Les figures 10.6 et 10.7 sont toutes deux des façons de remplir (comparées à la figure 10.5, par exemple) parce que toute procédure qui a besoin d'être informée sur la couleur d'une région peut, par inspection mécanique de cette région, extraire cette information. C'est un remplissage purement informationnel. Les systèmes sont entièrement arbitraires, cela va de soi. Nous pouvons aisément construire de nombreux systèmes équivalents de représentation – impliquant divers systèmes de codages, ou divers véhicules.

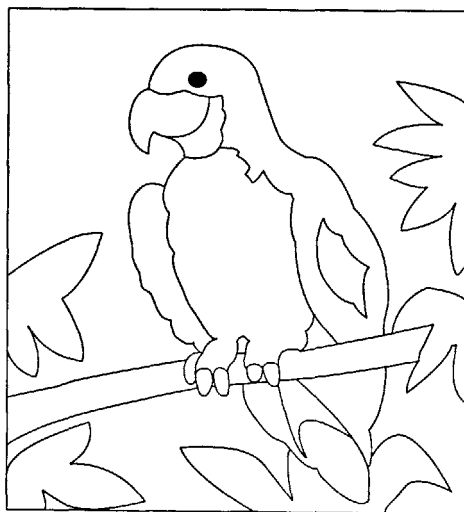
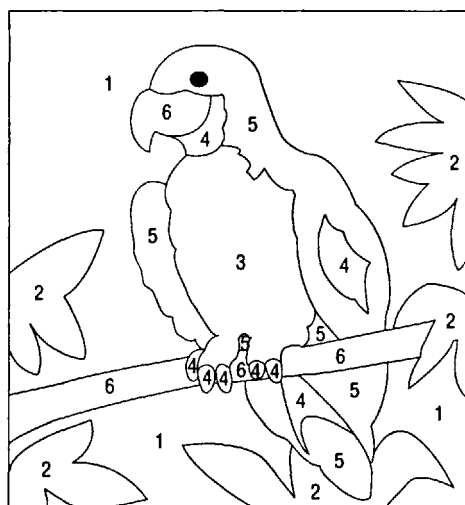


Figure 10.5



1 - Bleu	4 - Rouge
2 - Vert	5 - Pourpre
3 - Orange	6 - Jaune

Figure 10.6

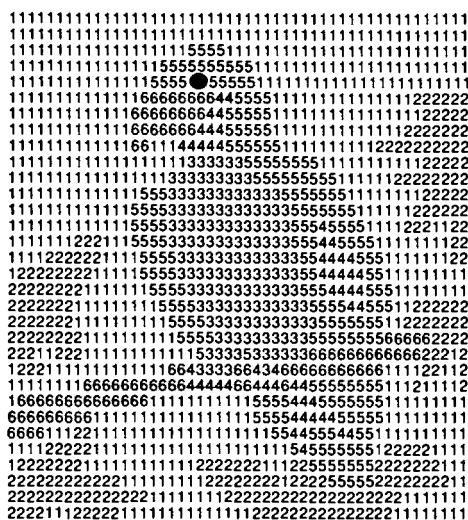


Figure 10.7

Si vous formez une image colorée sur votre ordinateur personnel en utilisant le programme pour PC « Paintbrush », l'écran que vous voyez est représenté dans la machine par une carte à bits dans le « butoir à cadres », analogue à la figure 10.7.

Mais quand vous vous stockez ensuite l'image sur une disquette, un algorithme de compression la traduit en quelque chose qui ressemble à la figure 10.6. Cet algorithme divise la surface en régions qui sont comme si elles étaient colorées, et stocke les limites des régions et leur nombre spécifique pour les couleurs dans un dossier d'« archivage »¹¹. Un dossier d'archivage est tout aussi précis qu'une carte à bits, mais il généralise sur les régions, et il donne une étiquette à chaque région une fois seulement, ce qui en fait un système de représentation plus efficace.

Une carte à bits, qui donne une étiquette explicite à chaque pixel, est une forme de ce que nous pouvons appeler une représentation *en gros continue* — « en gros » parce que cela dépend de la taille des pixels. Une carte à bits n'est pas, au sens littéral, une image, mais seulement un tableau de valeurs, une sorte de recette pour former une image. On peut stocker le tableau dans n'importe quel système qui préserve l'information portant sur les localisations. Une bande vidéo est un autre véhicule de représentation en gros continue, mais ce qu'elle stocke sur la bande ce ne sont pas des images, mais des recettes (de différents degrés de précision) pour former les images.

Une autre façon de stocker des images sur votre écran d'ordinateur consisterait à prendre une photographie en couleurs, et à stocker l'image sur une diapositive en 35 mm par exemple. Cela diffère beaucoup des autres systèmes, de façon évidente : il y a dans ce cas une teinte réelle, qui remplit littéralement une région de l'espace réel. Comme une carte à bits, c'est une représentation en gros continue des régions spatiales représentées (continue jusqu'au grain même du film), à une échelle suffisante, il finit par devenir granulaire ou couvert de pixels. Mais à la différence de la carte à bits, la couleur sert à représenter la couleur. Un négatif en couleurs utilise aussi la couleur pour représenter la couleur, mais avec une projection inverse.

Nous avons donc vu trois façons de « remplir » un espace avec de l'information sur des couleurs : la couleur par les nombres, comme dans la figure 10.6 ou sur un dossier d'archivage, la couleur-par-la-carte-à-bits, comme dans la figure 10.7 ou avec les butoirs à cadre ou les bandes vidéo, et les couleurs par-les-couleurs. La-couleur-par-les-nombres est en un sens une façon de « remplir » avec de l'information sur la couleur, mais elle devient efficace, comparée aux autres, précisément parce qu'elle

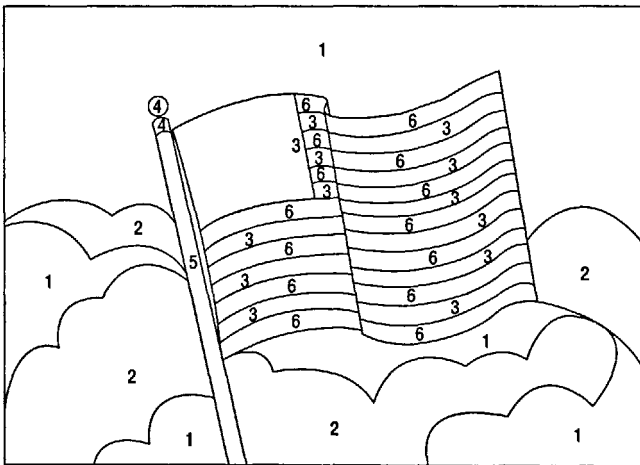
ne se soucie *pas* de remplir des valeurs explicitement pour chaque pixel. En quel sens, parmi ces trois possibles (si cela a un sens), le cerveau « remplit »-il la tâche aveugle ? Personne ne pense que le cerveau utilise des nombres dans des registres pour coder des couleurs, mais c'est une fausse piste. On peut considérer les nombres dans des registres comme représentant n'importe quel système de grandeurs, n'importe quel système de « vecteurs », qu'un cerveau pourrait employer comme un « code » pour des couleurs : cela pourrait être des fréquences de décharges de neurones, ou un système d'adresses ou de localisations dans des réseaux neuronaux, ou n'importe quel autre système de variations physiques dans le cerveau qu'on voudra. Les nombres dans des registres ont une propriété intéressante : ils préservent les relations entre des grandeurs physiques tout en restant neutres quant à des propriétés intrinsèques quelconques, en sorte qu'elles peuvent représenter n'importe quelle grandeur dans le cerveau qui « coderait » les couleurs. Bien que l'on puisse utiliser les nombres de façon parfaitement arbitraire, on peut aussi les utiliser de façon non arbitraire, pour refléter les relations structurales entre les couleurs découvertes. Le « solide de couleurs » dans lequel la teinte, la saturation et la luminosité sont les trois dimensions sous lesquelles *nos* couleurs varient ¹² est un espace logique qui s'adapte idéalement à un traitement numérique – n'importe quel traitement numérique reflétant des relations d'être – entre, les relations d'opposition et de complémentarité, et ainsi de suite, qui sont propres à la vision humaine. Plus nous en apprenons sur la façon dont le cerveau code la vision des couleurs, plus nous pourrions construire un modèle numérique puissant de la vision des couleurs humaine et réduire son caractère arbitraire.

Le problème, quand on dit que le cerveau « code » les couleurs en utilisant des intensités ou des grandeurs pour une chose ou pour une autre, est que cela suggère aux imprudents qu'il faut à un moment décoder ces codages, pour « revenir aux couleurs ». C'est l'une des voies – peut-être la mieux connue – pour revenir au figment : on imagine que le cerveau pourrait inconsciemment stocker son information encyclopédique sur la couleur dans un format qui ressemblerait à celui de la figure 10.8, mais qu'on aurait ensuite à s'arranger pour que la représentation soit « décodée » en « couleurs réelles » dans des circonstances particulières – comme en faisant marcher une vidéocassette, pour projeter une véritable couleur sur un écran. Il y a certainement une

différence de phénoménologie entre le fait de simplement rappeler la proposition que le drapeau est rouge, blanc et bleu, et le fait d'imaginer effectivement « la couleur » du drapeau et « voir » (dans l'œil de l'esprit) qu'il est rouge, blanc et bleu. Si ce contraste phénoménologique conduit certaines personnes à postuler l'existence d'un figment, il sera encore plus tentant de le faire dans le cas du phénomène de diffusion de la couleur néon (van Tuijl, 1975), dont on peut voir un exemple au dos de la couverture de ce livre.

La couleur rose que vous voyez remplir l'anneau défini par les lignes rouges n'est pas le résultat d'un barbouillage de la page en rose, ni de l'éparpillement de la lumière : il n'y a pas de rose sur votre image rétinienne, en plus des lignes rouges. Comment expliquer cette illusion ? Un circuit cérébral, spécialisé dans la forme, est faussement induit à distinguer une région circonscrite particulière : l'anneau avec ses « contours subjectifs ». Bien d'autres figures du même genre, comme la suivante, produisent des contours subjectifs.

Un autre circuit cérébral, spécialisé dans le traitement des couleurs mais plutôt pauvre pour celui des formes et des localisations, produit une discrimination de couleurs (disons le rose



1 - Bleu	4 - Doré
2 - Gris	5 - Noir
3 - Blanc	6 - Rouge

Figure 10.8

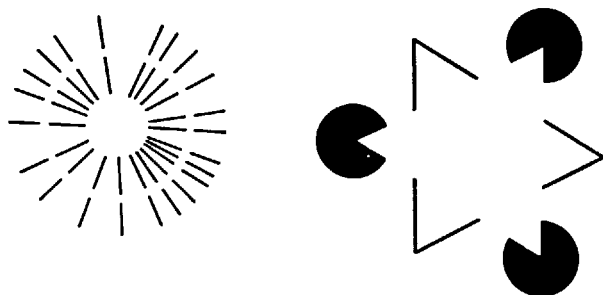


Figure 10.9

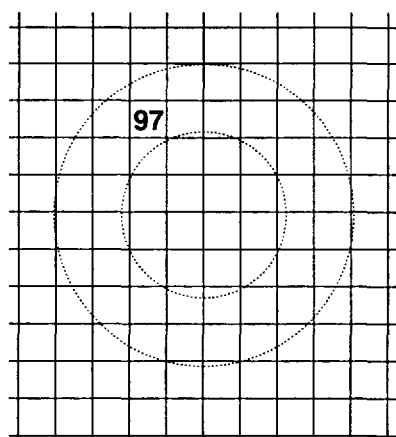


Figure 10.10

à 97) avec lequel on peut « étiqueter » quelque chose qui se trouve dans le voisinage, en sorte que l'étiquette s'attache (« ou se lie ») à l'ensemble de la région.

Pourquoi ces discriminations particulières devraient-elles intervenir dans ces conditions ? C'est encore un point débattu, mais la controverse concerne les mécanismes causaux qui conduisent à assigner à la région une mauvaise étiquette, et pas les « produits » ultérieurs (s'il y en a) du système visuel. Mais ne manque-t-il pas quelque chose ? Je me suis arrêté à une explication qui fournisse une région étiquetée par des nombres correspondants aux couleurs : cette recette de production d'une image colorée n'a-t-elle pas à être exécutée quelque part ? Le « rose à 97 » ne doit-il pas être « rempli » quelque part ? Après

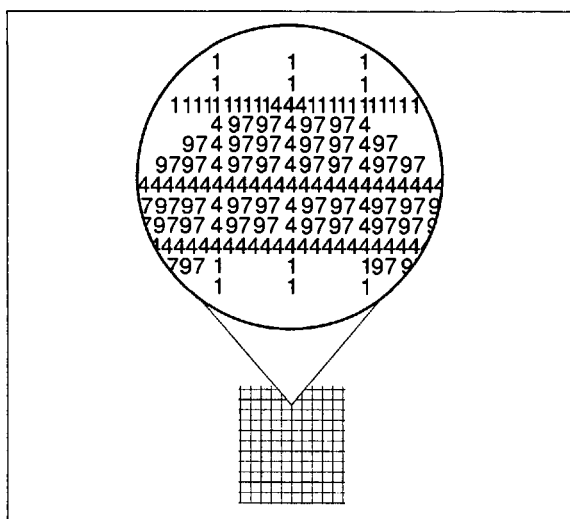


Figure 10.11

tout, vous pourriez être tenté d'insister sur le fait que vous voyez le rose ! Vous ne voyez certainement pas une région esquissée, avec un nombre au milieu. Le rose que vous voyez n'est pas dans le monde extérieur (ce n'est pas un pigment, une teinte, ou une « lumière colorée »), en sorte qu'il doive « être là » – c'est-à-dire un figment rose.

Nous devons faire attention de bien distinguer l'hypothèse du « figment rose » d'autres hypothèses rivales légitimes qu'on pourrait proposer pour contrer une explication qui s'arrêterait à la suggestion de la couleur-par-les-nombres. Par exemple, il pourrait se faire que quelque part dans le cerveau il existe une représentation en gros continue de régions colorées – une carte à bits – telle que « chaque pixel » dans la région ait à recevoir l'étiquette « couleur 97 », plus ou moins comme dans la figure 10.11.

C'est une possibilité empirique. Nous pourrions monter des expériences pour la confirmer ou l'infirmer. La question serait la suivante : y a-t-il un véhicule de représentation dans le cerveau dans lequel la valeur d'un certain paramètre variable (l'intensité ou tout ce qui peut coder la couleur) doit être propagée ou répliquée sur les pixels pertinents d'un tableau, ou n'y a-t-il qu'une seule étiquette de la région, sans qu'un « remplisse-

ment » ou un « étalement » soit nécessaire ? Quelles sortes d'expériences pourraient favoriser un tel modèle de l'effet de l'étalement de la couleur néon ? Il serait étonnant, par exemple, de constater qu'on peut montrer que la couleur se répand lentement sous certaines conditions – saignant à partir des lignes centrales pour atteindre graduellement les limites du contour subjectif¹³. Je ne veux pas préjuger de la question, parce que mon but principal en la posant est d'illustrer ma thèse : alors qu'il y a quantité de questions empiriques non résolues sur la façon dont le phénomène de l'étalement la couleur néon se répand dans le cerveau, aucune de ces questions n'implique qu'on réponde différemment à la question de savoir si le figment est engendré par un « décodage » du système de codage neuronal.

La question de savoir si le cerveau « remplit » d'une façon ou d'une autre n'est pas une question qui puisse par elle-même se régler en recourant à l'introspection, car, comme nous l'avons vu au chapitre 3, l'introspection nous fournit – au sujet aussi bien qu'à l'expérimentateur « extérieur » – seulement le contenu de la représentation, et pas les traits du véhicule de représentation lui-même. Pour savoir quelle est la nature du véhicule, il nous faut conduire d'autres expériences¹⁴. Mais pour certains phénomènes, nous pouvons déjà être *tout à fait* sûrs que le véhicule de représentation est une version de quelque chose d'efficace, comme les couleurs-par-les-nombres, et qui ne soit pas en gros continu, comme les cartes à bits.

Voyons comment le cerveau doit s'occuper du papier mural, par exemple. Supposez que vous entriez dans une pièce et que vous remarquiez que le papier mural est une suite continue de centaines de bateaux identiques, ou – en hommage à Andy Warhol – des portraits photographiques identiques de Marilyn Monroe. Pour identifier une image comme un portrait de Marilyn Monroe, il vous faut fover l'image : elle doit tomber sur les fovéas à haute résolution de vos yeux. Comme nous l'avons vu dans l'expérience de la carte à jouer au chapitre 2, votre vision *parafovéale* (servie par le reste de la rétine) n'a pas une très haute résolution : vous ne pouvez même pas identifier un valet de carreau tenu à bout de bras. Et pourtant nous savons que si vous deviez entrer dans une pièce dont les murs seraient tapissés de photos identiques de Marilyn Monroe, vous vous en apercevriez « à l'instant ». Vous verriez en une fraction de seconde qu'il y a « des quantités et des quantités de portraits identiques de Marilyn

Monroe ». Étant donné que vos yeux font des saccades quatre ou cinq fois par seconde au plus, vous pourriez fover seulement une ou deux Marilyn pendant la durée que cela vous prend de sauter jusqu'à la conclusion *et par suite de voir* des centaines de Marilyn Monroe identiques. Nous savons que la vision parafovéale *ne pourrait pas* distinguer Marilyn de diverses taches de couleur de forme marilynesque, mais néanmoins ce que vous voyez n'est *pas* du papier mural de Marilyn-au-milieu entourée de diverses taches indistinctes de forme marilynesque.

Est-il possible que le cerveau prenne l'une de ses images fovéales à haute résolution de Marilyn Monroe et la reproduise, comme s'il la photographiait, à travers une projection interne de la surface du mur ? Il n'y a que comme cela que les détails à haute résolution qui vous ont servi à identifier Marilyn ont pu « passer à l'arrière-plan », parce que la vision parafovéale n'est pas assez précise pour fournir par elle-même cet arrière-plan. Je suppose que c'est possible en principe, mais il est presque certain que le cerveau ne se préoccupe pas de faire *ce genre* de remplissement là ! Après avoir identifié une seule Marilyn, et sans avoir reçu d'information qui lui dise que les autres taches ne sont pas des Marilyn, il saute à la conclusion que le reste des taches ce sont des Marilyn, et il étiquète l'ensemble de la surface « d'autres Marilyn » sans représentation supplémentaire de Marilyn ¹⁵.

Bien entendu, ce n'est pas ainsi que les choses vous semblent se passer. Il vous semble voir effectivement des centaines de Marilyn identiques. Et en un sens, c'est ce qui se passe : il y a, évidemment, des centaines de Marilyn identiques là-bas sur le mur, et vous êtes en train de les voir. Ce qui ne se passe pas, en revanche, est qu'il y a des centaines de Marilyn identiques représentées dans votre cerveau. Votre cerveau se contente de représenter d'une certaine façon *qu'il y a* des centaines de Marilyn identiques, et aussi vive que soit votre impression que vous voyez le détail des choses, le détail est dans le monde, pas dans votre tête. Et aucun figment n'intervient dans la production de cette apparence, parce que l'apparence n'est pas produite du tout, même pas comme une carte à bits.

Nous pouvons donc maintenant répondre à la question que nous posions au sujet de la tache aveugle. Le cerveau n'a pas à « remplir » l'espace occupé par la tache aveugle, parce que la région dans laquelle tombe la tache aveugle est déjà étiquetée (par exemple « écossais » ou « Marilyn », ou simplement « autres

trucs du même genre »). Si le cerveau recevait des données qui contredisent celles d'une autre région, il abandonnerait la généralisation ou il l'ajusterait. Mais le fait de ne pas recevoir de données de la région de la tache aveugle n'est pas la même chose que de recevoir des données contradictoires. L'absence de confirmations empiriques venues de la région de la tache aveugle ne pose pas de problème au cerveau : puisqu'il n'a pas enregistré précédemment d'information de cette lacune dans la rétine, il n'a pas développé d'agences épistémiquement affamées qui réclameraient de la nourriture venue de cette région. Parmi tous les homoncles de la vision, il n'y en a aucun dont le rôle consisterait à coordonner de l'information venue de cette région de l'œil, et c'est pourquoi quand aucune information ne vient de ces sources, personne ne se plaint. L'aire en question est simplement négligée. En d'autres termes, toutes les personnes douées d'une vision normale « souffrent » d'une petite « anosognosie ». Nous ne nous rendons pas compte de ce « déficit » – du fait que nous ne recevons pas d'information visuelle de nos taches aveugles. (On trouve une bonne analyse de l'anosognosie chez Mc Glynn et Schacter, 1989.)

La tache aveugle est un trou spatial, mais il peut y avoir aussi des trous temporels. Les plus petits sont les lacunes qui interviennent quand nos yeux font leur mouvement en saccades. Nous ne faisons pas attention à ces lacunes, mais elles n'ont pas à être remplies *parce que* nous sommes conçus pour ne pas y faire attention. Les analogues temporels des scotomes pourraient être les « absences » qui se produisent durant les crises épileptiques. Celui qui souffre les remarque, mais seulement par inférence : ils ne peuvent pas plus « voir les bords » que vous ne pouvez voir les bords de votre tache aveugle, mais ils peuvent être frappés, rétrospectivement, par des discontinuités dans les événements dont ils ont eu l'expérience.

Le vice fondamental inhérent à l'idée de « remplissement » est qu'elle suggère que le cerveau est en train de fournir quelque chose alors qu'il est en train d'en ignorer une. Et c'est ce qui conduit des penseurs au demeurant fort subtils à faire des erreurs graves, qu'Edelman illustre parfaitement : « L'un des traits les plus frappants de la conscience est sa continuité » (1989, p. 119). C'est parfaitement faux. L'un des traits les plus frappants de la conscience est sa *discontinuité*, comme le montre la tache aveugle, les lacunes dans les saccades, pour ne prendre que les exemples

les plus simples. La discontinuité de la conscience est frappante en raison de la continuité *apparente* de la conscience. Neumann (1990) fait remarquer que la conscience peut en général être un phénomène lacunaire, et que tant que les bords temporels des lacunes ne sont pas positivement perçus, on ne pourra pas ressentir le caractère lacunaire du « flux » de la conscience lui-même. Comme le dit Minsky, « rien ne peut *sembler* être chaotique sauf ce qui est *représenté* comme étant chaotique. Paradoxalement, notre sens de la continuité vient de notre merveilleuse *insensibilité* à la plupart des formes de changement, plutôt que d'une authentique capacité à les percevoir » (1985, p. 257).

La négligence comme perte pathologique d'appétit épistémique

La devise du cerveau quand il a affaire à la tâche aveugle devrait être : ne me posez pas de questions et je ne vous dirai pas de mensonges. Comme nous l'avons vu au chapitre premier, tant que le cerveau apaise les appétits épistémiques qu'il peut avoir, il n'a pas besoin de faire quoi que ce soit. Mais que se passe-t-il quand il y a beaucoup moins d'appétit épistémique qu'il ne devrait y en avoir ? Ce sont les pathologies de la négligence.

L'une des formes les plus familières de négligence est l'*hémiparésie*, dans laquelle un côté du corps, habituellement le côté gauche, est entièrement négligé, en raison d'une lésion cérébrale du côté opposé. Non seulement le côté gauche du corps, mais aussi le côté gauche du voisinage immédiat sont négligés. Si un groupe de personnes se tiennent autour du lit du patient victime de négligence sur son côté gauche, il ne regardera que les gens qui se tiennent sur sa droite ; si on lui demande de compter les gens qui sont dans la pièce, il tendra à ignorer les gens qui se trouvent à sa gauche, et si quelqu'un sur sa gauche essaie d'attirer son attention, cela échouera de façon typique. Et pourtant on peut montrer que les organes sensoriels du patient sont toujours réceptifs, qu'ils analysent, et répondent de diverses façons aux stimuli qui surviennent sur sa gauche. Que peut-il bien se passer dans la tête du patient ? Le « côté gauche de l'espace phénoménal est-il un blanc » ? Ou faut-il dire que l'« œil de l'esprit » du patient ne parvient pas à saisir les données que le cerveau lui fournit sur la gauche de... la scène du Théâtre Cartésien ?

Il y a une explication simple, qui ne recourt pas à des représentations internes douées de propriétés curieuses, mais en termes de *négligence* – au sens politique du terme ! Daniel Patrick Moynihan est célèbre pour avoir défendu l'idée que les relations raciales en Amérique se résoudraient d'elles-mêmes si on les traitait par une « légère négligence » – si Washington et le reste du pays les ignoraient tout simplement pendant un temps. Je ne pense pas que ce soit une bonne idée, mais Moynihan avait raison sur un point : il y a certainement des cas où on a besoin d'une légère négligence – comme celui de notre traitement du problème de la tache aveugle.

Il n'y a pas d'homoncules, comme je l'ai dit, qui sont supposés « faire attention » à l'information venue de la part du champ visuel qui est couverte par la tache aveugle. C'est pourquoi quand rien ne se passe, il n'y a personne pour protester. La différence entre nous et les victimes de négligence pathologique ou d'autres formes d'anosognosie tient peut-être au fait que certains protestataires ont été tués. Cette théorie a été avancée, en des termes moins pittoresques, par le neuropsychologue Marcel Kinsbourne (1980), qui appelle les protestataires internes des « analyseurs corticaux ». Selon le modèle que nous avons développé, la négligence se décrirait comme une perte de puissance politique par certains partis de démons dans le cerveau, due, dans de nombreux cas (mais pas dans tous), à la mort ou à la suppression de leur Représentant. Ces démons sont toujours actifs, essayant de mener leurs diverses activités et y réussissant même à l'occasion, mais ils ne peuvent plus gagner dans certaines compétitions contre des coalitions mieux organisées.

Selon ce modèle, la négligence légère de nos taches aveugles se fond presque imperceptiblement dans les diverses négligences semi-dysfonctionnelles dont nous sommes tous victimes, pour rejoindre finalement les négligences les plus bizarres qu'étudient les neurophysiologistes. Par exemple, je souffre moi-même de formes plutôt communes de négligence. La moins sérieuse, bien qu'elle soit quelquefois embarrassante, est ma *typo*-négligence. Je suis pathologiquement incapable de relever des erreurs typographiques dans les épreuves de mes écrits quand je les lis, et ce n'est que grâce à des exercices laborieux de concentration et d'attention que je peux vaincre ce défaut. Ce n'est pas, comme l'a suggéré Baars, que mon cerveau « remplisse » en mettant l'écriture correcte ; il n'a pas à « remplir » parce qu'il

n'a pas ordinairement à faire assez attention à cela pour remarquer les erreurs ; son attention est captée par d'autres traits des mots sur la page. Une autre de mes incapacités mineures est la *négligence des copies d'examens d'étudiants*. C'est extraordinaire de voir à quel point la perspective de laver le sol de la cuisine, de changer le papier de mon armoire, ou de vérifier mes comptes sur mon carnet de chèques devient attirante quand j'ai une pile de copies d'examens à noter sur mon bureau. Ce trait, l'intérêt accru pour d'autres tâches, est particulièrement évident dans l'hémi-négligence : en première approximation, plus une chose se trouve sur la droite, plus elle est digne d'attention pour un patient victime d'hémi-négligence. Mais la forme la plus sérieuse de négligence dont je souffre est peut-être ma *négligence financière*. J'aime si peu vérifier mes comptes sur mon carnet de chèques, en fait, qu'une autre tâche possible qui m'est très pénible, telle que noter des copies d'examens, peut forcer mon attention à la tâche en question. Cette négligence a de sérieuses conséquences pour mon train de vie, des conséquences que je peux aisément être conduit à vérifier, mais en dépit de l'insuccès alarmant de cet appel à ma rationalité foncière, je m'arrange pour persister dans ma négligence, jusqu'à ce que des mesures particulièrement drastiques d'auto-manipulation interviennent.

Ce n'est pas que je *ne puisse pas voir* mon carnet de chèques. C'est plutôt que *je n'ai pas envie de le regarder*. Et bien que dans des moments calmes, réfléchis, comme celui-ci, je puisse rendre compte de tout cela (pourvu que je ne sois pas *profondément* anosognosique quant à ma propre incapacité), dans le cours usuel des événements une partie de ce que je ne remarque pas est ma propre négligence financière. Bref, c'est une forme atténuée d'anosognosie. De ce point de vue, la seule qui soit étonnante dans les formes bizarres de négligences qu'étudient les neuropsychologues est la question des frontières. Imaginez que quelqu'un néglige tout ce qu'il y a *sur la gauche* (Bisiach et Luzzatti, 1978 ; Bisiach, 1988, Bisiach et Vallar, 1988 ; Calvanio, Petrone et Levine, 1987). Ou imaginez quelqu'un qui aurait perdu la vision des couleurs *mais qui ne s'en plaindrait pas* (Geschwind et Fussilo, 1966). Ou même, imaginez quelqu'un qui serait devenu aveugle, mais qui n'aurait pas remarqué ce déficit profond – le syndrome d'Anton ou la dénégation de cécité (Anton, 1899 ; Mc Glynn et Schacter, 1989, p. 154-158).

On peut expliquer facilement ces cas avec la théorie des Versions Multiples de la conscience, parce que le Témoin Central a été remplacé par des coalitions de spécialistes dont les appétits épistémiques particuliers ne peuvent pas être immédiatement adoptés par d'autres agents s'ils sont effacés ou en vacances ¹⁶. Quand ces appétits épistémiques s'évanouissent, ils s'évanouissent sans laisser de traces, en laissant le champ libre pour d'autres coalitions, d'autres agents avec d'autres agendas.

Mais le principe qui permet l'analyse de la négligence nous fournit aussi un scénario possible permettant d'expliquer les « qualia visuels manquants » du virtuose de la vision aveugle que nous avons imaginés. J'ai suggéré qu'il était possible que s'il se plaignait de l'absence de qualia, il *pourrait* simplement relever la pauvreté relative de l'information qu'il reçoit présentement de sa vision et la décrire mal. J'ai aussi spéculé que si nous pouvions d'une façon ou d'une autre augmenter le « taux de bauds » de son système de réception d'information, une partie, sinon l'intégralité, de la distance qui sépare sa forme de vision de la vision normale se trouverait réduite. Nous pouvons voir à présent qu'il existe une autre façon, plus simple, de réduire cette distance : il suffit d'affaiblir son appétit épistémique, d'obstruer sa curiosité visuelle d'une certaine façon. Après tout, si dans le syndrome d'Anton un sujet peut être totalement aveugle et pourtant ne pas le réaliser, un petit peu de négligence placée de façon stratégique pourrait transformer notre sujet à vision aveugle qui se plaint de la perte de ses qualia visuels en un sujet qui ne se plaint pas et qui déclare que sa vision est à présent restaurée et sans défaut. Il pourrait sembler que nous puissions en apprendre plus, mais le pourrions-nous ? Quelque chose manquerait-il à ce genre de personne ? Il n'y a pas de *figment* dans la vision normale : ce ne peut donc pas être un figment qui manque. Quoi d'autre alors ?

La présence virtuelle

« Nous avons le sens de la réalité quand toute question portant sur nos systèmes visuels reçoit une réponse si rapide qu'il semble que ces réponses étaient déjà là. »

Marvin Minsky, 1985, p. 257.

Répetons-le une fois de plus, l'absence de représentation n'est pas la même chose que la représentation d'une absence. Mais c'est difficile à croire. Notre conviction d'être en quelque façon *directement en contact (acquainted)* avec des propriétés ou des traits spéciaux de notre expérience est l'une des intuitions les plus puissantes qui s'opposent à toute tentative de fournir une bonne théorie de la conscience. J'ai mis cette idée en pièces, et essayé de miner son autorité, mais je ne suis pas au bout de mes peines. Otto revient à la charge :

Ce que tu dis des Marilyn sur le papier mural est en fait une défense en sous-main du dualisme. Tu soutiens de façon très convaincante qu'il n'y a pas des centaines de Marilyn à haute résolution dans le cerveau, et tu conclus qu'il n'y en a pas où que ce soit ! Mais je soutiens que puisque *ce que je vois* ce sont des centaines de Marilyn à haute résolution, il s'ensuit que, *puisque*, comme tu le soutiens, elles ne sont nulle part dans mon cerveau, elles doivent être quelque part ailleurs – dans mon esprit non physique !

Les centaines de Marilyn sur le papier mural semblent être présentes dans votre expérience, elles semblent être dans votre esprit, et pas seulement sur le mur. Mais puisque, comme nous le savons, votre regard peut se déplacer en une fraction de seconde pour extraire de l'information de n'importe quelle partie de votre environnement visuel, pourquoi votre cerveau se préoccuperait-il donc d'importer toutes ces Marilyn ? Pourquoi ne pas laisser le monde les stocker, sans que cela lui coûte quoi que ce soit, jusqu'à ce qu'on en ait besoin ?

Comparons le cerveau à une bibliothèque. Certaines bibliothèques universitaires sont des chambres de stockage gigantesques, contenant à l'intérieur de leurs murs des millions de

livres, tous aisément accessibles dans les rayonnages. D'autres bibliothèques tiennent moins de livres à la disposition immédiate des lecteurs, mais ont des systèmes d'accès généreux et efficaces, et achètent tous les livres que les utilisateurs réclament, ou les empruntent à d'autres bibliothèques, par l'intermédiaire d'un système de prêt inter-bibliothèques. Si vous ne conservez pas les livres sur les lieux, les délais d'accès sont plus grands, mais pas tant que cela. Nous pouvons imaginer un système électronique de prêt inter-bibliothèques (utilisant la télécopie ou des fichiers informatiques) qui pourrait obtenir un livre du monde extérieur plus rapidement que ne le pourrait le coureur le plus rapide qui retirerait le livre des rayonnages. Un informaticien pourrait dire des livres obtenus par un tel système qu'ils étaient « virtuellement présents » dans la bibliothèque dès le départ, ou que la « collection virtuelle » de la bibliothèque était des centaines ou des milliers de fois plus vaste que sa collection d'exemplaires effectivement présents sur les rayonnages.

À présent, comment pourrions-nous, en tant qu'utilisateurs de nos bibliothèques cérébrales, savoir lesquelles, parmi les choses que nous en extrayons, étaient *déjà là* dès le début, et que notre cerveau a envoyé chercher, par des incursions rapides de recherche d'information en direction du monde extérieur ? Si l'on fait avec soin des expériences, selon la méthode hétérophénoménologique, on pourra répondre à cette question, mais l'introspection par elle-même ne pourra pas le dire. Mais cela ne nous empêche pas de penser que nous ne pouvons pas le dire. En l'absence de preuves empiriques dans un sens ou dans un autre, nous avons naturellement tendance à nous précipiter pour conclure qu'il y a quelque chose de plus. C'est ce que j'ai appelé le Piège Introspectif (Dennett, 1969, p. 139-140) et c'est ce que Minsky appelle l'illusion d'Immanence : « Chaque fois que vous pouvez répondre à une question sans qu'il y ait de délai perceptible, il vous semble que la réponse était déjà active dans votre esprit » (Minsky, 1985, p. 155).

Le système de prêt inter-bibliothèques est une analogie utile, mais incomplète, parce que votre cerveau n'a pas seulement la capacité d'acquérir de l'information de n'importe quelle source externe qui puisse vous intéresser ; il a aussi littéralement des millions de sentinelles qui sont continuellement en train de contempler une portion du monde extérieur, prêtes à sonner l'alarme et à *attirer* votre attention sur tout ce qui peut se passer

de nouveau et de pertinent dans le monde. Dans le mécanisme de la vision, cette tâche incombe aux bâtons et cônes parafovéaux des rétines, et aux agents neuronaux qui se trouvent à l'intérieur de ces sentinelles, qui sont spécialisés dans la détection du changement et du mouvement. Si l'un de ces agents tire la sonnette d'alarme – « Changement dans mon secteur ! » – cela déclenche presque instantanément une saccade, en conduisant la fovéa à se porter sur la région devenue objet d'intérêt, en sorte que la nouveauté peut être localisée, identifiée, traitée. Le système des sentinelles est si fiable qu'il est difficile d'introduire un changement dans le monde visible sans que l'ensemble du système visuel en soit informé. Mais avec l'aide de techniques avancées, il arrive que l'on puisse échapper aux sentinelles, et les résultats sont surprenants.

Quand vos yeux se meuvent par saccades, les contractions musculaires qui conduisent les globes oculaires à tourner sont des actions *balistiques* : vos points de fixation sont des missiles *non* guidés dont les trajectoires au décollage déterminent où et quand elles atteindront à leur point d'impact une nouvelle cible. Par exemple, si vous lisez un texte sur l'écran d'un ordinateur, vos yeux sauteront sur quelques mots à chaque saccade, d'autant plus loin et plus vite que vous êtes un bon lecteur. Quel effet cela ferait-il si un magicien, une sorte de Malin Génie cartésien à petite échelle, pouvait changer le monde pendant les quelques millisecondes pendant lesquelles vos yeux se pointent vers leurs nouvelles destinations ? À notre grand étonnement, un ordinateur équipé d'un détecteur de vision peut détecter et analyser le décollage dans les quelques premières millisecondes d'une saccade, calculer où sera le point d'impact, et, *avant la fin d'une saccade*, peut effacer le mot sur l'écran au point d'impact et le remplacer par un nouveau mot de même longueur. Que voyez-vous ? Juste le nouveau mot, sans avoir en quoi que ce soit le sentiment que quoi que ce soit ait été changé. Quand vous utilisez le texte sur l'écran, il vous semble totalement aussi stable que si les mots étaient gravés dans du marbre. Mais pour une autre personne qui lirait le texte par-dessus votre épaule (et dont les saccades seraient produites par un autre tambour) l'écran serait traversé de changements.

L'effet est puissant. Quand j'ai été confronté pour la première fois à une expérience avec un détecteur de vision, et que j'ai constaté combien les sujets pouvaient être inattentifs (en appa-

rence) aux changements qui papillotaient sur l'écran, je demandai si je ne pouvais pas être sujet moi aussi, et voulus m'en assurer par moi-même. On m'assit devant le dispositif, et on immobilisa ma tête en me faisant mordre une « barre à mordre ». Cela facilite le travail au détecteur de vision, qui envoie un rayon imperceptible de lumière sur les lentilles de l'œil du sujet, et analyse le retour pour détecter tout mouvement de l'œil. Pendant que j'attendais que les expérimentateurs enclenchent le dispositif, je lus le texte sur l'écran, j'attendis, et attendis, impatient que l'essai commence. Je m'impatientai. « Pourquoi ne le mettez-vous pas en marche ? », demandai-je. « Il est en marche », répondirent-ils.

Comme tous les changements sur l'écran se produisent pendant des saccades, vos sentinelles ne produisent aucune sonnerie d'alarme. Jusqu'à récemment, on appelait ce phénomène « suppression saccadique ». On voulait dire par là que le cerveau doit d'une façon ou d'une autre fermer l'entrée venue des yeux pendant des saccades, puisque personne ne peut relever les changements qui se produisent dans le champ visuel pendant les saccades, et bien entendu, personne ne se plaint d'éprouver des changements vertigineux et alarmants. Mais une expérience habile avec un détecteur de vision (Brooks et al., 1990) a montré que si un stimulus – tel qu'un mot ou une lettre de l'alphabet – était mû en synchronie avec la saccade, gardant le rythme avec l'« ombre » de la fovéa quand elle s'en va vers son nouveau point d'atterrissage, il est immédiatement vu et identifié par le sujet. L'entrée venue de l'œil n'est pas interrompue dans son voyage vers le cerveau, mais dans les conditions normales il est inusable – tout va trop vite pour qu'on s'en aperçoive – c'est pourquoi le cerveau le traite par une négligence légère. Si toutes vos sentinelles envoient leurs alarmes en même temps, ce qu'il y a de mieux à faire c'est de les ignorer toutes.

Dans la situation expérimentale dans laquelle je me trouvais, les mots sur l'écran étaient effacés et remplacés pendant mes saccades. Si votre vision parafovéale ne peut pas discriminer un mot à l'impact avant que vous ne lui envoyiez une saccade, à partir de moment où vous l'atteignez et l'identifiez, il ne peut pas y avoir d'enregistrement préalable ou de souvenir de ce mot dans votre cerveau avec lequel vous puissiez le comparer. On ne peut pas remarquer le déplacement, parce que l'information logiquement requise pour pouvoir le remarquer n'est tout simple-

ment *pas là*. Bien sûr, il *vous semble* quand vous lisez cette page que tous les mots sur la ligne sont en quelque sorte présents dans votre conscience (dans l'arrière-plan) avant même que vous n'y fassiez spécifiquement attention, mais c'est une illusion. Ils ne sont que virtuellement présents.

Il existe *une certaine* information sur les mots environnants dans votre cerveau, bien entendu – suffisamment pour avoir servi de guide et d'instigateur pour la saccade la plus récente, par exemple. Quelle quantité d'information exactement se trouve-t-elle déjà *là* ? Les expériences avec les détecteurs de vision et d'autres dispositifs peuvent déterminer les limites de ce que vous pouvez remarquer, et par là déterminer les limites de ce qui est *présent* dans votre esprit (voir par exemple Pollatsek, Rayner et Collins, 1984 ; Morris, Rayner et Pollatsek, 1990). Insister sur le fait, comme Otto était tenté de le faire, que ce qui n'est pas *là* dans le cerveau doit néanmoins être *là* dans l'esprit parce qu'il *semble* certainement s'y trouver est vain. Car, comme nous venons de le voir, cela ne serait pas « *là* » en un sens quelconque qui pourrait changer la nature *des expériences propres à Otto*, voire sa capacité à passer des tests, à presser des boutons, et ainsi de suite.

Voir c'est croire : un dialogue avec Otto

C'est le moment où notre critique, Otto, nous demande de faire le point, parce qu'il est sûr d'avoir été berné à un moment donné. Je vais le faire dialoguer avec moi, en espérant qu'il rend justice à bon nombre de nos doutes, sinon à tous. Otto prend la parole.

Il me semble que tu as nié l'existence des phénomènes les plus indubitablement réels qui soient : les apparences réelles que même Descartes dans ses *Méditations* ne pouvait soumettre au doute.

– En un sens, tu as raison : c'est ce dont je nie l'existence. Revenons au phénomène de la couleur néon qui se répand. Il semble qu'il y ait un rayonnement rosâtre sur la couverture du livre.

– Sans aucun doute.

– Mais il n'y a pas d'anneau rosâtre qui rayonne. Pas vraiment.

– Certes. Mais il semble certainement qu'il y en a un !

– Certes.

– Où se trouve-t-il, alors ?

- Où est quoi ?
- L'anneau rosâtre qui rayonne.
- Il n'y en a pas ; je pensais que tu reconnaîtrais précisément cela.
- Eh bien oui, il n'y a pas d'anneau rosâtre sur la page, mais il ne fait pas de doute qu'il semble qu'il y en ait un.
- Certes. Il semble qu'il y ait un anneau rosâtre qui rayonne.
- Concentrons-nous donc sur *cet* anneau.
- Lequel ?
- Celui-là même qui *semble* être présent.
- Il n'y a pas de chose telle qu'un anneau rosâtre qui semble simplement exister.
- Écoute, je ne *dis* pas simplement qu'il semble y avoir un anneau rosâtre qui rayonne ; *il me semble réellement* qu'il y a un anneau rosâtre qui rayonne !
- Je m'empresse d'acquiescer. Je ne t'accuserais jamais d'insincérité ! Tu veux vraiment dire ce que tu veux dire quand tu dis qu'il semble y avoir un anneau rosâtre qui rayonne.
- Écoute. Je ne me contente pas de vouloir le dire. Je ne *pense* même pas qu'il semble y avoir un anneau rosâtre qui rayonne ; il semble *réellement* y avoir un anneau rosâtre qui rayonne !
- Ça y est. Tu es tombé dans un piège, et beaucoup d'autres gens avec toi. Tu sembles penser qu'il y a une différence entre le fait de penser (juger, décider, croire du fond du cœur) que quelque chose te semble être rose et le fait que quelque chose te *semble réellement* être rose. Mais il n'y a pas de différence. Il n'existe pas de phénomène, tel que le sembler réel qui vienne s'ajouter au phénomène de juger comme ceci ou comme cela que quelque chose est le cas.
- Rappelle-toi le papier mural avec les Marilyns. Le mur est en fait couvert des Marilyns à haute résolution. De plus, c'est ainsi qu'il te semble être ! Il te semble que le mur est couvert de Marilyns à haute résolution. Par chance pour toi, ton appareil visuel t'a conduit à une croyance vraie quant à un trait de ton environnement. Mais il y a des quantités d'*apparences réelles* de Marilyns représentées dans ton cerveau – ou ton esprit. Il n'y a pas de véhicule qui *rend* le détail du papier mural, qui le *restitue* à ton Témoin intérieur. La seule chose qui se passe est qu'il te semble qu'il y a là des quantités de Marilyns à haute résolution (et cette fois tu as raison – elles sont vraiment là). Dans les autres cas, tu peux avoir tort ; il peut te sembler – dans le

phénomène phi pour les couleurs – qu'un point unique s'est déplacé vers la droite, qu'il a changé de couleur, alors qu'en fait il y avait simplement deux points colorés lumineux qui s'allumaient. Le fait que cela te soit apparu ainsi ne requiert pas un *rendu* dans le cerveau, pas plus que les jugements du cerveau quant aux couleurs, une fois atteints, n'ont besoin d'être ensuite *décodés* quelque part.

– Mais alors que se passe-t-il donc quand il me semble qu'il y a un anneau rosâtre ? Quelle explication positive peux-tu en donner avec ta théorie ? Tu sembles être terriblement évasif à ce sujet.

– Je suppose que tu as raison. Il est temps de nous mettre au net et de présenter l'explication positive, mais j'avoue qu'il va me falloir le faire en commençant par une caricature que je réviserai ensuite. Il semble que je ne parvienne pas à découvrir une façon plus directe de l'exposer.

– Je m'en suis bien aperçu. Vas-y.

– Suppose qu'il y *ait* un Signifieur Central. Mais suppose qu'au lieu de siéger dans le Théâtre Cartésien en train de contempler la Présentation, le Signifieur Central soit assis dans le noir et qu'il ait des pressentiments – il réalise tout soudain qu'il y a quelque chose de rose là-devant, tout comme tu pourrais soudain réaliser qu'il y a quelqu'un debout devant toi.

– Que sont exactement des pressentiments ? De quoi sont-ils faits ?

– Bonne question, à laquelle je dois répondre d'abord évasivement, sous forme de caricature. Ces pressentiments sont des propositions que le Signifieur Central s'exclame à lui-même dans son propre langage, le mentalais. Sa vie consiste donc en une suite de *jugements*, qui sont des phrases du mentalais, exprimant une proposition, puis une autre, à une vitesse extraordinaire. Il y a certaines de ces phrases qu'il décide de publier, en traduction française.

Cette théorie a l'avantage de se débarrasser du figment, la projection dans l'espace phénoménal, le remplissement de tous les blancs sur l'écran du Théâtre, mais elle a toujours un Signifieur Central, et le Langage de la Pensée. Par conséquent, révisons la théorie. Tout d'abord, débarrassons-nous du Signifieur Central en distribuant tous ses jugements dans le temps et dans l'espace dans le cerveau – chaque acte de discrimination ou de discernement ou de fixation de contenu se produit quelque part, mais

il n'y a pas un Discerneur qui ferait tout le travail. Et ensuite, débarrassons-nous du Langage de la Pensée : le contenu des jugements n'a pas besoin d'être exprimé sous une forme « propositionnelle » – c'est une erreur, une forme de projection erronée trop confiante des catégories du langage sur les activités du cerveau.

– Les pressentiments sont donc comme des actes de langage, à cette nuance près qu'il n'y a pas d'Acteur ni de Langage !

– Eh bien oui. Ce qu'il y a, réellement, ce sont seulement divers événements de fixation de contenu qui se produisent à différents moments dans le cerveau. Ce ne sont les actes de langage de personne, et par conséquent ils n'ont pas besoin de faire partie d'un langage, mais ils sont plutôt *comme* des actes de langage ; ils ont un contenu, et ils ont bien comme effet d'informer divers processus dotés de ce contenu. Nous avons considéré des versions plus détaillées de cette idée aux chapitres 5 à 10. Certaines de ces fixations de contenu ont des effets supplémentaires, qui conduisent finalement à l'énoncé de phrases – dans le langage naturel – soit publiques soit purement internes. Et c'est ainsi qu'un texte hétérophénoménologique se crée. Quand il est interprété, il se crée l'illusion bénigne d'un Auteur. Cela suffit pour produire l'*hétérophénoménologie*.

– Mais qu'en est-il de la phénoménologie *effective* ?

Il n'y a rien de tel. Rappelle-toi notre discussion de l'interprétation de la fiction. Quand nous rencontrons un roman qui est de l'autobiographie à peine voilée, nous découvrons que nous pouvons projeter les événements fictionnels sur de nombreux événements réels de la vie de l'auteur, et c'est pourquoi en un sens un peu forcé le roman *porte* sur ces événements réels. L'auteur peut ne pas le réaliser, mais, en ce sens forcé, c'est vrai ; ces événements sont ce sur quoi porte le texte, parce que ce sont les événements réels qui expliquent *pourquoi* ce texte s'est trouvé créé.

– Mais sur quoi porte le texte, en un sens *non* forcé ?

– Sur rien. C'est de la fiction. Il *semble* porter sur divers personnages, lieux et événements fictionnels, mais ces événements ne se sont jamais produits ; il ne porte pas *réellement* sur quoi que ce soit.

– Mais quand je lis un roman, ces événements fictionnels s'animent ! Quelque chose se passe en moi ; je *visualise* les événements. L'acte de lire et d'interpréter un texte tel qu'un roman

crée certaines choses nouvelles dans mon imagination : des images des personnages qui accomplissent les actions décrites. Après tout, quand nous allons voir la version filmique d'un roman que nous avons lu, nous pensons souvent : « Ce n'était pas du tout comme cela que j'avais imaginé cette femme ! »

– Tout à fait d'accord. Dans « Fearing Fictions » le philosophe Kendall Walton (1978) soutient que ces actes d'imagination de la part d'un interprète viennent s'ajouter au texte pratiquement de la même façon que les images qu'on trouve dans les versions illustrées des romans, « se combinant avec le roman pour former un monde (fictionnel, hétérophénoménologique) plus “ vaste ” » (p. 17). Ces additions sont parfaitement réelles, mais elles sont juste du « texte » en plus – elles ne sont pas faites de figment, mais de jugement. La phénoménologie ne revient pas à autre chose.

– Mais il semble y en avoir !

– Exactement ! *Il semble y avoir de la phénoménologie*. C'est un fait que l'hétérophénoménologue concède avec enthousiasme. Mais de ce fait indéniable, universellement attesté, *il ne s'ensuit pas* qu'il y ait *réellement* de la phénoménologie. C'est là le point crucial.

– Et tu nies par conséquent que la conscience soit un *plein* ?

– Oui, certes. C'est l'une des choses que je nie. La conscience est pleine de lacunes et clairsemée, et elle ne contient pas la moitié de ce que les gens croient qu'elle contient !

– Mais, mais....

– Mais la conscience semble bien être un plein ?

– Oui !

– Je l'admets ; elle semble être un plein : cela semble même être « un fait frappant » relatif à la conscience : elle est continue, comme le dit Edelman. Mais...

– Je sais, je sais : il ne s'ensuit pas du fait qu'il *semble* y avoir un plein qu'il y en ait *vraiment* un.

– Voilà ! Tu as saisi.

– Mais je vois un autre problème dans ce palais des glaces que tu appelles une théorie. Tu dis que tout se passe seulement *comme* s'il existait un Signifieur Central, *comme* s'il y avait un unique Auteur ; *comme* s'il y avait un lieu où tout se rejoint ! Je ne comprends pas toute cette histoire de *comme si* !

– Peut-être une autre expérience de pensée rendra-t-elle tout cela plus clair. Imagine que nous nous rendions dans une autre

planète et que nous découvririons que là-bas les savants ont une théorie charmante : toute chose physique contient une âme, et toute âme aime toutes les autres âmes. De ce fait, les choses tendent à se mouvoir en direction les unes des autres, poussées par l'amour mutuel que se portent leurs âmes intérieures. Nous pouvons aussi supposer que ces savants ont construit des systèmes très précis de position d'âme, si bien que, une fois qu'ils ont déterminé la localisation précise dans l'espace de l'âme d'une chose, ils peuvent répondre à des questions portant sur sa stabilité (« elle tombera parce que son âme est trop haute »), sur sa vibration (« si vous posez un contrepoids sur le côté de ce volant, qui a une âme assez vaste, il va se mettre à trembler ») et sur bien d'autres points techniques.

Ce que nous pourrions leur dire, bien entendu, c'est qu'ils sont tombés sur le concept de centre de gravité (ou plus précisément de centre de masse), et qu'ils le traitent simplement un peu trop cérémonieusement. Nous leur disons qu'ils peuvent très bien continuer à parler et à penser comme ils le faisaient – mais qu'il leur faut seulement laisser tomber une partie de leur bagage métaphysique inutile. Il y a une interprétation plus simple, plus austère (et bien plus satisfaisante) des faits mêmes qu'ils cherchent à comprendre avec leur physique spirituelle. Ils nous demandent : Y a-t-il des âmes ? Évidemment, répondons-nous – si ce n'est que ce sont des *abstracta*, des abstractions mathématiques plutôt que des pépites d'une substance mystérieuse. Ce sont des fictions merveilleusement utiles. Tout se passe *comme si* tout objet attirait tout autre objet en concentrant toute son énergie (*oomph*) gravitationnelle en un point unique – c'est bien plus facile que de calculer le comportement de systèmes utilisant cette fiction fondée que cela ne le serait de descendre jusqu'aux détails peu attrayants – tout point attire tout autre point.

– Je me sens *comme si* on venait de me fouiller les poches.

– Eh bien, ne dis pas que je ne t'avais pas prévenu. Tu ne peux pas t'attendre à ce que la conscience soit *précisément* ce que tu souhaitais qu'elle soit. Qui plus est, à quoi renonces-tu ?

– Seulement à mon âme.

– Pas en un sens cohérent, sensé. Tout ce à quoi tu renonces c'est une pépite d'un minéral spécial qui ne pourrait pas être bien spécial de toute manière. Pourquoi aurais-tu une meilleure opinion de toi-même s'il se révélait que tu es une sorte de perle

mentale dans l'huître de ton cerveau ? Qu'y aurait-il de si spécial dans le fait d'être une perle mentale ?

– Une perle mentale pourrait être immortelle. Pas un cerveau.

– L'idée que le Soi – ou l'Âme – n'est en fait qu'une abstraction frappe beaucoup de gens comme n'étant qu'une idée négative, un refus plutôt que quelque chose de positif. Mais en fait elle a beaucoup d'avantages, y compris – si cela compte pour toi – une version de l'immortalité potentielle quelque peu plus robuste que quoi que ce soit qu'on puisse trouver dans les notions traditionnelles de l'âme. Mais pour cela il nous faudra attendre le chapitre 12. Tout d'abord il nous faut régler leur compte aux qualia, qui *continuent* toujours à exercer une sorte de fascination sur notre imagination.

NOTES

1. Tim Shallice, dans *From Neuropsychology to Mental Structure* (1988), donne une discussion très à jour et très bien argumentée du raisonnement qu'impliquent ces expériences sur la nature. Plusieurs livres récents donnent de bonnes vulgarisations de certains de ces cas fascinants : Howard Gardner, *The Shattered Mind* (1975), et Oliver Sacks, *L'homme qui prenait sa femme pour un chapeau* (1985).

2. Notons que les détails portant sur les lésions cérébrales ne prouvent en eux-mêmes rien (en l'absence des dénégations) : ce n'est qu'en établissant une correspondance entre la lésion cérébrale et les comptes rendus (crédibles) et les données comportementales que nous pouvons obtenir des hypothèses sur les parties du cerveau qui sont essentielles pour tel ou tel phénomène conscient.

3. S'il n'y avait pas la confirmation fournie par les scanners cérébraux révélant la lésion corticale, nous éprouverions aussi un grand scepticisme quant à l'authenticité des scotomes des sujets victimes de vision aveugle. Voir par exemple Campion, Latto et Smith (1983), et Weizkranz (1988).

4. Le philosophe Colin Mc Ginn (1991), dit à propos d'un patient victime de vision aveugle : « Comportementalement, elle peut fonctionner comme une personne qui voit ; phénoménologiquement, elle se considère comme aveugle » (p. 111). C'est tout simplement faux ; elle ne peut pas du tout fonctionner comportementalement comme une personne dotée de la vue. Mc Ginn tempère ensuite son affirmation frappante : « De plus, soyons naïfs pendant un instant : les patients victimes de vision aveugle ne *semblent-ils* pas tout à fait avoir une expérience visuelle quand ils font leurs discriminations surprenantes ?... Ils ne semblent pas être comme le sont les gens ordinairement quand ils n'éprouvent *aucune* expérience » (p. 112). À nouveau, c'est faux. Ils semblent en fait *ne pas* avoir d'expériences visuelle, *parce qu'il faut leur donner des indices*. S'il ne fallait pas leur donner des indices, ils sembleraient en fait avoir des expériences visuelles – à tel point que nous ne croirions pas leurs dénégations !

5. « S'il pouvait écouter sa propre réponse galvanique sur sa peau, il serait en meilleure forme » – parole de Larry Weizkranz commentant l'un de ses patients victime de vision aveugle (ZMF, Bielefeld, mai 1990).

6. Ma réponse à cette question se trouve dans mon livre *La Stratégie de l'interprète* (1987, tr. 1990).

7. Son identification du tampon est-il un *effet* ultérieur au fait qu'elle en est devenue consciente, ou est-ce une *cause* antérieure au fait qu'elle en soit devenue consciente ? Là est la question – orwellienne ou stalinesque – que le modèle des Versions Multiples nous pose.

8. Dans les conditions normales, la localisation (« le détecter ») et l'identification vont de pair : détecter la chose à identifier est une précondition de son identification. Mais cette coïncidence normale masque un fait frappant : la machinerie qui permet l'identification et la machinerie qui fait la localisation sont dans une large mesure indépendantes dans le cerveau, localisées dans différentes régions du cortex (Mishkin, Ungerleider et Macko, 1983) et par conséquent capables d'être éteintes immédiatement. Il existe quelques rares pathologies dans lesquelles le sujet peut aisément identifier *ce qu'il* voit sans être capable de le localiser dans l'espace personnel, et il y a des pathologies homologues dans lesquelles les sujets peuvent localiser un stimulus visuel – pointer en sa direction, par exemple – et pourtant être très peu capables d'identifier l'objet, en dépit du fait qu'ils aient par ailleurs une vision tout à fait normale. Anne Treisman (1988, Treisman et Gelade, 1980 ; Treisman et Sato, 1990, Treisman et Souther, 1985) a fait une série importante d'expériences qui tendent à montrer selon elle que *voir* doit être distingué d'*identifier*. Quand quelque chose est vu, selon son modèle, le cerveau établit une « occurrence » (*token*) de l'objet. Les « occurrences » sont des « représentations séparées, épisodiques » – et leur création est le préambule de leur identification ultérieure, quelque chose qui s'accomplit par recherche dans la mémoire sémantique en utilisant un processus du type de ceux que modélisent les systèmes de production. Mais une occurrence n'a pas besoin d'être définie par une localisation définie dans l'espace personnel, si je comprends son modèle, et pour cette raison il n'est pas hors de question que les sujets qui sont dans l'état dans lequel se trouvait Marinette (avant de trouver le tampon) puissent faire mieux que s'ils procèdent au hasard si on les pousse à faire une conjecture forcée pour dire si le tampon est dans leur champ visuel ou pas. Pour des expériences sur ces questions, voir Pollatsek, Rayner et Henderson (1990).

9. Par exemple, le temps de latence de la réponse pour certaines de ces tâches perceptuelles, même chez des sujets entraînés, est très long (Bach-y-Rita, p. 103). Cela par soi-même démontre que le flux d'information dans la vision prothétique est extrêmement paresseux, comparé à celui de la vision normale.

10. Le « taux de bauds » est le terme standard désignant le flux d'information digitale (cela veut dire, approximativement : bits par seconde). Par exemple, si votre ordinateur communique par des lignes téléphoniques avec d'autres ordinateurs, il peut transmettre ses suites de bits à 1 200 bauds ou à 2 400 bauds ou à une vitesse bien supérieure. Cela prend un taux de bauds approximativement quatre fois plus rapide de transmettre des images animées de haute résolution en temps réel – ce qui montre clairement qu'une image vaut plus qu'un millier de mots. Les signaux de télévision usuels sont analogiques, comme un enregistrement de phonographe, pas digitaux, comme un disque compact, en sorte que son flux d'information se mesure comme une *largeur de bande*, plutôt que comme taux de bauds. Ce terme date d'avant l'invention des ordinateurs ; le code Baudot, du nom de son inventeur (comme l'était le code Morse), était le code standard télégraphique adopté en 1880, et le taux de bauds était le nombre d'éléments de code transmis par seconde. En utilisant le « taux de bauds » plutôt que « largeur de bande », je ne veux pas impliquer par là que la manipulation de l'information se conçoit mieux en termes digitaux.

11. Il y a d'autres sortes d'algorithmes de compression qui ne reposent pas sur une décomposition de l'image en régions colorées de la même manière de cette façon, mais je ne les discuterai pas.

12. D'autres créatures utilisent des solides de couleurs différents – ou des hypersolides ! Nous sommes des « trichromates » : nous avons trois types différents de transducteurs photopigmentés de cellules dans les cônes de nos rétines.

D'autres espèces, comme les pigeons, sont des tétrachromates ; leur espace subjectif de couleurs devrait être représenté, numériquement, comme un hyperspace à quatre dimensions. D'autres espèces sont des dichromates, toutes celles dont les discriminations de couleurs pourraient être projetées sur un seul plan bidimensionnel. (Notez que « noir-et-blanc » est juste un schème de représentations unidimensionnel, où tous les gris possibles représentent autant de distances différentes sur une ligne entre 0 et 1.) On trouvera des réflexions sur les implications de cette incommensurabilité des systèmes de couleur chez Hardin (1988) et Thomson, Palacios et Varela (sous presse).

13. Comme pour répondre à cette suggestion V.S. Ramachandran et R.L. Gregory (article soumis pour publication) viennent juste de faire certaines expériences qu'ils appellent (de façon trompeuse, à mon sens) des scotomes induits artificiellement, qui montrent bien qu'il y a un remplissement graduel par des textures et des détails. Il y a une différence fondamentale entre leurs conditions expérimentales et celles que j'ai décrites ; dans leurs expériences, il y a compétition entre deux sources d'information, et l'une d'elles se trouve dépassée (graduellement). Le phénomène du remplissement spatial graduel par des textures est une découverte importante, mais elle ne nous permet pas d'aller plus loin qu'un modèle dans l'esprit de la figure 10.11. Et il faut résoudre d'autres questions au sujet de ces expériences avant que l'on puisse établir définitivement comment on doit les interpréter.

14. Par exemple, les expériences initiales de Roger Shepard sur la rotation mentale de diagrammes représentant des cubes montraient qu'il *semblait* certainement aux sujets qu'ils faisaient des représentations en gros continues des formes qu'ils imaginaient, mais qu'il fallait d'autres expériences, sondant les propriétés temporelles effectives des représentations sous-jacentes, pour fournir une confirmation partielle de l'hypothèse selon laquelle ils étaient bien en train de faire ce qu'il leur semblait qu'ils faisaient. (Voir Shepard et Cooper, 1982.)

15. Dans l'Appendice B, je suggérerai « quelques expériences avec du papier mural » qui pourront permettre de tester cette thèse empirique.

16. Pour un autre point de vue, voir Bisiach et alii (1986) et Mc Glynn et Schacter (1989), dont les modèles de l'anosognosie sont similaires, mais reposent sur la « boîteologie » de systèmes séparés (en particulier Mc Glynn et Schacter, qui postulent l'existence d'un *système d'attention consciente*, qui prenne des entrées venues de modules).

Chapitre 11

Les qualia disqualifiés

Un nouveau fil de cerf-volant

« Si on le jette dans une fosse causale, un quale passera tout simplement au travers. »

Ivan Fox, 1989.

Quand votre cerf-volant se met à faire des nœuds, vous pouvez en principe les défaire, surtout si vous êtes patient et analytique. Mais, passé un certain stade, les principes vacillent et c'est le côté pratique qui l'emporte. Il faut tout simplement renoncer à défaire certains nœuds. Et aller acheter un nouveau fil de cerf-volant. En vérité, c'est finalement bien meilleur marché que la peine que cela vous coûterait de réparer l'ancien et votre cerf-volant pourra plus tôt se remettre à voler. À mon avis, c'est exactement quelque chose de ce genre qui se passe en philosophie avec la question des « qualia ». On a affaire là à un sac de nœuds de plus en plus embrouillé et compliqué d'expériences de pensées, de jargon, de jeux de mots, d'allusions bizarres à des réfutations supposées, à des résultats » reçus « qu'il faudrait renvoyer à l'expéditeur, ainsi qu'à maintes voies de garage et autres pertes de temps. Il y a certains guêpiers qu'il vaut mieux éviter, aussi ne vais-je pas m'engager dans une visite guidée de tout ce qui a pu être écrit sur le sujet, même si j'ai pu faire mon profit, çà et là, d'idées perspicaces et ingénieuses (Shoemaker, 1975, 1981, 1989, White, 1986, Kitcher, 1979, Harman, 1988, Fox, 1989). Par le passé, j'ai essayé de défaire les nœuds de cette question (Dennett,

1988), mais je pense à présent qu'il vaut mieux essayer de reprendre les choses presque à zéro.

Il n'est guère difficile de voir les nœuds dans lesquels les philosophes se sont empêtrés au sujet des qualia. Ils sont partis du point dont partirait toute personne douée de bon sens : avec les intuitions les plus fortes et les plus claires qu'ils pouvaient avoir sur leur esprit. Malheureusement, ces intuitions forment un cercle clos de doctrines autosuffisantes, qui enferment leur imagination dans le Théâtre Cartésien. Même si les philosophes ont découvert les paradoxes inhérents à ce cercle clos d'idées – d'où tous ces textes sur les qualia – ils n'ont pas trouvé *une autre vision globale* à laquelle s'accrocher, et c'est ainsi que, faisant confiance à leurs intuitions encore fortes, ils se sont laissé ramener dans la prison paradoxale. C'est pourquoi tous les écrits portant sur les qualia sont de plus en plus alambiqués, au lieu de se dissoudre en un accord.

Mais nous avons désormais mis en place une autre vision, et c'est justement le Modèle des Versions Multiples. En l'utilisant, nous pouvons proposer une analyse assez différente des questions qui se posent. Nous marquerons un temps de réflexion aux quatrième et cinquième sections de ce chapitre pour la comparer aux visions que, je l'espère, elle remplacera.

On trouve, dans un excellent livre d'introduction sur le cerveau, le passage suivant :

« “ La couleur ” comme telle n'existe pas dans le monde ; elle n'existe que dans l'œil et le cerveau de celui qui regarde. Les objets reflètent de nombreuses longueurs d'onde différentes, mais ces longueurs d'onde elles-mêmes n'ont pas de couleur » (Ornstein et Thompson, 1984).

C'est une manière heureuse d'exprimer la sagesse commune, mais on notera qu'à strictement et littéralement parler, ce ne peut être ce que veulent dire les auteurs, et cela ne peut être vrai. La couleur, disent-ils, n'existe pas « dans le monde » ; elle n'existe que « dans l'œil et dans le cerveau du spectateur ». Mais l'œil et le cerveau du spectateur sont dans le monde, et sont autant partie intégrante du monde physique que les objets que voit le spectateur. Et comme ces objets, l'œil et le cerveau sont colorés. Les yeux peuvent être bleus, bruns ou verts, et même le cerveau n'est pas *uniquement* fait de matière grise (et blanche) : en plus de la *substantia nigra* (la substance noire), il y a le *locus*

ceruleus (l'endroit bleu). Mais il va de soi que les couleurs qui sont « dans l'œil et le cerveau du spectateur » en ce sens ne sont pas ce dont parlent les auteurs. Qu'est-ce qui fait penser qu'il y a de la couleur en un quelque autre sens ?

La science moderne – c'est du moins ce qu'on dit – a ôté toute couleur au monde physique, pour la remplacer par des radiations électromagnétiques incolores de différentes longueurs d'onde, rebondissant sur des surfaces qui reflètent et absorbent diversement ces radiations. Il peut sembler que la couleur est *là-bas au-dehors*, mais il n'en est rien. Elle est *ici au-dedans* – dans « l'œil et le cerveau de l'observateur. » (Si les auteurs du passage n'étaient pas d'aussi bons matérialistes, ils auraient probablement dit qu'elle n'est que dans l'*esprit* de l'observateur, s'épargnant l'interprétation ridicule que nous venons de rejeter, mais se créant des problèmes encore pires.) Cela dit, s'il n'y a aucune *fiction* intérieure qui puisse être colorée en un sens spécifique, subjectif, intérieur à l'esprit, phénoménal, c'est alors tout bonnement à une disparition des couleurs qu'on assiste ! Il doit bien y avoir *quelque chose* comme les couleurs que nous connaissons et aimons, que nous mélangeons et accordons. Où, mais où, peuvent-elles bien être ?

Telle est la bonne vieille devinette philosophique qu'il nous faut à présent affronter. Au XVII^e siècle, le philosophe John Locke (et avant lui, le savant Robert Boyle) donnait le nom de *qualités secondes* à ces propriétés que sont les couleurs, les odeurs, les goûts et les sons. On les distinguait des qualités premières : taille, figure, mouvement, nombre et solidité. Les qualités secondes n'étaient pas elles-mêmes des choses-dans-l'esprit, mais plutôt les pouvoirs des choses dans le monde (grâce à leurs qualités premières particulières) de produire ou de provoquer certaines choses dans l'esprit de l'observateur normal. (Qu'advenait-il alors s'il n'y avait aucun observateur alentour ? C'est l'éternelle énigme populaire de l'arbre qui tombe dans la forêt. Émet-il un son ? On laisse la réponse aux bons soins du lecteur.) La manière dont Locke a défini les qualités secondes est devenue partie intégrante de l'interprétation standard de la science qu'ont les profanes, et elle a ses vertus, mais elle a aussi ses gages : les choses produites dans l'esprit. La qualité seconde, rouge, par exemple, était pour Locke la propriété dispositionnelle ou le pouvoir de certaines surfaces d'objets physiques, grâce à leurs aspects microscopiques structurels, de produire en nous l'*idée de rouge* chaque fois que

la lumière réfractait ces surfaces dans vos yeux. Le pouvoir qui existe dans un objet externe est apparemment assez clair. Mais en quoi consiste exactement une idée de rouge ? Est-elle en un sens *colorée*, ainsi qu'une belle robe faite de bleu ? Ou porte-t-elle simplement, comme une belle discussion de pourpre, *sur* une couleur, sans être elle-même le moins du monde colorée ? Voilà qui ouvre des possibilités, mais comment une idée pourrait-elle simplement porter *sur* une couleur (par exemple, la couleur rouge) si rien, où que ce soit, n'est rouge ?

Qu'est-ce que le rouge, après tout ? En quoi consistent les couleurs ? La couleur a toujours été l'exemple favori des philosophes et je suivrai pour le moment la tradition. Le problème majeur que pose la tradition apparaît clairement dans l'analyse philosophique de Wilfrid Sellars (1963, 1981), qui a distingué les propriétés dispositionnelles des objets (les qualités secondes de Locke) de ce qu'il a appelé les propriétés occurrentes. Un cube de glace rose dans la glacière, toutes lumières éteintes, a la qualité seconde rose, mais il n'existe aucune instance de la propriété de *rose occurrent* tant qu'un observateur n'ouvre pas la porte pour regarder. Le rose occurrent est-il une propriété de quelque chose qui se trouve dans le cerveau ou de quelque chose qui relève du « monde extérieur » ? Dans un cas comme dans l'autre, souligne Sellars, le rose occurrent est une propriété « homogène » de quelque chose de réel. Ce qu'il voulait partiellement dénier en insistant sur l'homogénéité, c'était l'hypothèse que le rose occurrent soit quelque chose comme *l'activité neuronale d'intensité 97 dans la région 75* du cerveau. Il voulait aussi dénier que le monde subjectif de la phénoménologie de la couleur se ramène à quelque chose d'aussi incolore que *le fait de juger* que telle ou telle chose soit, ou semble être rose. Par exemple, l'acte consistant à rappeler à l'œil de votre esprit la couleur d'une banane mûre et de juger qu'il s'agit de la couleur jaune ne suffirait pas à soi seul à faire surgir une instance de jaune occurrent (Sellars, 1981 ; Dennett, 1981). Ce serait simplement juger que quelque chose est jaune, phénomène qui est en soi aussi dénué de jaune occurrent que le serait un poème sur des bananes.

Sellars est allé jusqu'à prétendre qu'une révolution était nécessaire dans les sciences de la nature qui accorderait droit de cité au rose occurrent et autres choses du même genre. Peu de philosophes l'ont suivi sur cette voie radicale, mais le philosophe Michael Lockwood en a fait ressurgir récemment une version

(1989). D'autres philosophes, tels que Thomas Nagel, ont supposé que même une science qui aurait subi une révolution serait incapable de traiter de telles propriétés :

« Les aspects subjectifs des processus mentaux conscients – à l'inverse de leurs causes et effets physiques – ne peuvent être saisis par la forme purifiée de pensée qui, elle, convient pour traiter du monde physique sous-jacent aux apparences » (1986, p. 15).

Les philosophes ont adopté différents noms pour les choses qui sont dans le spectateur (ou propriétés du spectateur) censées accorder droit de cité aux couleurs et au reste des propriétés bannies du monde « extérieur » par les conquêtes de la physique : « sensations brutes », « perceptions sensibles », « qualités phénoménales », « propriétés intrinsèques des expériences conscientes », « contenu qualitatif des états mentaux », et bien sûr, « qualia », terme que j'utiliserai. Il y a de subtiles différences dans la manière de définir ces termes, mais je n'en tiendrai pas compte. Au chapitre précédent, j'ai eu l'air de refuser l'existence de telles propriétés et, pour une fois, la réalité est bien conforme aux apparences. Je refuse en effet l'existence de telles propriétés. Mais (et revoilà ce fameux thème) j'accorde bien volontiers qu'il existe apparemment des qualia.

Il existe apparemment des qualia parce qu'il semble bien que la science nous ait montré qu'il ne peut y avoir de couleurs là-bas au-dehors, et qu'elles doivent donc se trouver ici au-dedans. Il semble en outre que ce qui se trouve ici au-dedans ne peut pas se ramener simplement aux jugements que nous faisons quand les choses nous paraissent colorées. Mais ce raisonnement est confus. Ce que la science nous a réellement montré c'est simplement que les propriétés des objets, qui réfléchissent la lumière, mettent les êtres dans divers états distinctifs, répartis çà et là dans leur cerveau, et sous-tendant une foule de dispositions innées et d'habitudes acquises de complexité diverse. Et en quoi consistent *leurs* propriétés ? Ici, nous pouvons jouer une seconde fois la carte de Locke : ces états distinctifs du cerveau de l'observateur ont diverses propriétés « primaires » (leurs propriétés mécaniques dues à leurs connexions, aux états d'excitation de leurs éléments, etc.), et en vertu de ces propriétés primaires, ils ont diverses propriétés secondaires, purement dispositionnelles. Chez les êtres humains doués de langage, par exemple, ces états distinctifs disposent souvent en définitive les êtres à

exprimer des jugements verbaux qui font allusion à la « couleur » de diverses choses. La sémantique de ces énoncés fait clairement ressortir ce que sont censées être les couleurs : des propriétés réfléchissantes de la surface des objets, ou des volumes transparents (le cube de glace rose, le rayon de lumière oxhydrique). Et c'est tout simplement ce qu'elles sont en fait – bien que ce ne soit pas une mince affaire de dire de quel genre de propriétés réfléchissantes il s'agit en fait (pour des raisons que nous analyserons dans la section suivante).

Est-ce que nos états internes distinctifs n'ont pas *aussi* des propriétés « intrinsèques » spéciales, les propriétés subjectives, privées, ineffables qui constituent *la manière dont les choses nous apparaissent à l'œil* (à l'oreille, au nez, etc.) ? Ces autres propriétés seraient les qualia, et avant de jeter un coup d'œil aux arguments que les philosophes ont proposés pour essayer de *prouver* l'existence de ces autres propriétés, nous essaierons d'emblée de supprimer les raisons qu'on pourrait avoir de croire en ces propriétés, en trouvant d'autres explications possibles des phénomènes qui les rendent apparemment indispensables. On verra alors aisément par où pèchent systématiquement les preuves invoquées.

Selon cette autre conception, les couleurs *sont* des propriétés qui sont finalement « là-bas au-dehors ». Au lieu des « idées de rouge » chères à Locke nous avons (chez les êtres humains normaux) des états distinctifs qui ont le contenu *rouge*. Un exemple fera ressortir très clairement en quoi consistent ces états distinctifs – et ce qui est plus important, en quoi ils ne consistent pas. Nous pouvons comparer les couleurs des choses du monde en les mettant côte à côte et en les regardant, pour voir à quel jugement nous parvenons, mais nous pouvons aussi comparer les couleurs des choses tout simplement en nous les rappelant ou en les imaginant « dans notre esprit ». Est-ce que le rouge standard des rayures du drapeau américain sont du même rouge, ou d'un rouge plus foncé ou plus clair ou plus vif ou plus ou moins orange que le rouge standard de l'habit du Père Noël (ou d'une boîte aux lettres anglaise ou de l'étoile rouge soviétique) ? (Si aucun de ces deux exemples standards ne vous est présent à la mémoire, essayez-en deux autres, une carte bleue Visa par exemple et un ciel bleu, ou la feutrine verte de la table de billard et la pomme verte Granny Smith, ou le jaune du citron et le jaune du beurre.) Nous sommes capables de faire de telles comparaisons

« dans l'œil de notre esprit » et ce faisant, de faire en quelque sorte apparaître quelque chose en nous qui libère de l'information de notre mémoire et nous permet de comparer, dans une expérience consciente, les couleurs des objets standards telles que nous nous les rappelons (ou en tout cas comme il nous semble nous les rappeler). Certains d'entre nous sont plus doués que d'autres pour cela, et bon nombre d'entre nous n'ont pas grande confiance dans les jugements auxquels nous parvenons dans de telles circonstances. C'est pourquoi nous ramenons à la maison des échantillons de peinture, ou apportons des échantillons de tissus au magasin de couleurs, pour pouvoir étaler côte à côte dans le monde extérieur des instances des deux couleurs que nous souhaitons comparer.

Quand nous faisons ces comparaisons « dans l'œil de notre esprit », que se passe-t-il, selon moi ? Quelque chose de strictement analogue à ce qui se passerait dans une machine – un robot – qui pourrait faire elle aussi de telles comparaisons. Souvenez-vous au chapitre 9 du CADAVEUGLE Série I *Vorsetzer* (celui dont l'appareil photo pouvait être dirigé vers l'écran CAD). Supposez que nous placions en face de lui une photo en couleur du Père Noël et que nous lui demandions si le rouge de la photo est plus profond que le rouge du drapeau américain (quelque chose qu'il a déjà emmagasiné dans sa mémoire). Voici ce qu'il ferait : libérer de sa mémoire sa représentation de Old Glory et situer les rayures « rouges » (dans son diagramme, elles portent l'étiquette « rouge h 163 »). Puis il comparerait ce rouge au rouge de l'habit du Père Noël sur la photo en face de son appareil, qui se trouve être transduite par son système graphique de couleurs sous le nom de rouge h 172. Il comparerait les deux rouges *en soustrayant 163 de 172 et en obtenant 9*, qui indiquerait, par exemple, dans son interprétation, que le rouge du Père Noël (*lui*) paraît un peu plus profond et plus riche que le rouge du drapeau américain.

Cette histoire est délibérément hypersimple, pour mieux faire ressortir ce que je veux faire comprendre : il est évident que le CADAVEUGLE de Série I n'utilise pas de fiction pour faire comprendre la représentation tirée de sa mémoire (ou de sa perception courante), *mais nous non plus*. Le CADAVEUGLE de Série II ne sait probablement pas comment il compare les couleurs de quelque chose qu'il voit avec quelque chose dont il se souvient *et nous non plus*. Le CADAVEUGLE de Série I a – admettons – un

espace chromatique plutôt simple, appauvri, doté de bien peu des associations ou des tendances constitutives de l'espace chromatique personnel d'un être humain, mais hormis cette énorme différence en complexité dispositionnelle, il n'y a aucune différence importante. Je pourrais même dire qu'il n'y a pas de différence *qualitative* entre la manière dont le CADAVEUGLE s'acquitte de sa tâche et la nôtre. Les états distinctifs de CADAVEUGLE de Série I ont un contenu tout autant et pour les mêmes raisons que les états cérébraux distinctifs par lesquels j'ai remplacé les idées de Locke. *Certes*, le CADAVEUGLE de Série I n'a aucun qualia (ou du moins, c'est sur ce point que je vais faire bondir les amoureux des qualia). Ma comparaison implique donc bien qu'à mes yeux, nous n'avons pas davantage de qualia. Le *genre* de différence que les gens croient pouvoir trouver entre une machine et un individu vivant une vie d'homme (souvenez-vous de la machine à goûter que nous avons imaginée au chapitre 2) est une différence que je refuse catégoriquement : il n'y a aucune différence de ce genre. Ce n'est qu'une impression.

Pourquoi y a-t-il des couleurs ?

Quand Otto, au chapitre 10, jugeait qu'il semblait y avoir un anneau rosâtre qui rayonne, quel était le contenu de son jugement ? Si, comme je l'ai souligné, son jugement ne portait pas sur un quale, une propriété d'apparence d'anneau « phénoménal » (produit d'une fiction), sur quoi portait-il donc ? Quelle propriété se voyait-il tenté d'attribuer (faussement) à quelque chose de situé au-dehors dans le monde ?

Nombreux sont ceux qui ont remarqué combien il est étrangement difficile de dire tout simplement de quelles propriétés des choses du monde peuvent bien relever les couleurs. L'idée simple et attirante – qu'on trouve encore dans bien des discussions élémentaires – est que chaque couleur peut être associée à une longueur d'onde lumineuse unique, et que la propriété d'être rouge consiste simplement en la propriété de réfléchir toute la lumière de longueur d'onde rouge et d'absorber toutes les autres longueurs d'onde. Mais on a montré depuis un certain temps que c'était faux. Des surfaces dotées de différentes propriétés fondamentales de réflexion peuvent être vues comme étant de même couleur, et une même surface dans différentes conditions

d'éclairage peut être vue comme étant de différentes couleurs. Les longueurs d'onde de la lumière qui pénètrent dans l'œil ne sont qu'indirectement reliées aux couleurs dont nous voyons que sont dotés les objets (voir Gouras, 1984 ; Hilbert, 1987 ; et Hardin, 1988, pour une revue détaillée et diversifiée). Pour ceux qui avaient espéré en une manière simple et élégante de recevoir les bénéfices de la note prometteuse de Locke sur les pouvoirs dispositionnels des surfaces, la situation pourrait difficilement être plus sinistre. Certains (Hilbert par exemple, 1987) ont décidé de donner un ancrage objectif à la couleur en déclarant que c'est une propriété relativement manifeste des objets extérieurs, telle que la propriété de « réflectance spectrale de surface » ; ayant fait ce choix, ils sont alors tenus de conclure que la vision normale de la couleur nous présente souvent des illusions, puisque les phénomènes constants que nous percevons cadrent si mal avec les phénomènes constants de réflectance spectrale de surface que mesurent les instruments scientifiques. D'autres ont conclu qu'il vaut mieux considérer de manière subjective les propriétés chromatiques comme des propriétés à définir strictement en termes de systèmes d'états cérébraux de l'observateur, ignorant la variété et la confusion du monde qui provoque ces états : « Les objets colorés sont des illusions, mais des illusions non fondées. Nous sommes normalement dans des états perceptuels chromatiques, et ce sont des états neurologiques » (Hardin, 1988, p. 111. Voir Thompson, Palacios et Varela (à paraître) pour une discussion critique de ces opérations et d'autres arguments en faveur de l'option ici retenue.)

Ce qui est indiscutable c'est qu'il n'y a pas de propriété simple, non disjonctive, des surfaces telles que toutes les surfaces dotées de cette propriété et elles seules soient rouges (au sens de la qualité seconde lockéenne). C'est un fait de prime abord troublant, voire déprimant, puisqu'il semble suggérer que notre saisie perceptuelle du monde est bien pire que nous ne l'avions pensé – que nous vivons dans quelque chose comme un monde de rêve, ou que nous sommes les victimes d'illusions de masse. Notre vision des couleurs ne nous donne pas accès à des propriétés simples des objets, même si elle en a l'air. Pourquoi ?

Simple malchance ? Projet mal conçu ? Pas du tout ! Il y a une autre perspective bien plus éclairante, que nous pouvons adopter sur la couleur, qui m'a été montrée pour la première fois par la philosophe des neurosciences Kathleen Akins (1989,

1990) ¹. Parfois de nouvelles propriétés apparaissent pour telle raison. Un exemple particulièrement utile est celui du fameux cas de Julius et Ethel Rosenberg, qui furent condamnés et exécutés en 1953 pour avoir espionné le projet atomique américain pour le compte de l'Union soviétique. Il apparut à leur procès qu'à un moment donné ils avaient improvisé un astucieux système de mot de passe : on déchirait en deux une boîte en carton, et on apportait les morceaux à deux individus qui devaient prendre bien soin de se reconnaître. Chaque lambeau devenait pratiquement une pièce à conviction et un « détecteur unique » de celui qui allait avec ; lors d'une rencontre ultérieure chaque partie pouvait produire son morceau, et si les morceaux s'adaptaient parfaitement l'un à l'autre, tout allait bien. Pourquoi ce système marche-t-il ? Parce que déchirer le carton en deux produit un bord d'une complexité informationnelle qu'il serait pratiquement impossible de reproduire par construction délibérée. (Notons que si l'on découpait la boîte de carton avec un cutter et un rasoir, on manquerait totalement le but recherché.) Le bord cisaillé particulier d'un morceau devient un procédé *pratiquement* unique de recognition de schéma pour celui qui va avec ; c'est un appareil ou un transducteur permettant de détecter la propriété formelle M, où M est uniquement instancié par le morceau qui va avec lui.

En d'autres termes, la propriété formelle M et le détecteur de propriété-M qui la détecte étaient faits l'un pour l'autre. Il n'y aurait aucune raison pour que l'autre existe, ait été créée, en l'absence de l'autre. Et cela vaut aussi pour les couleurs et la vision des couleurs : elles étaient faites l'une pour l'autre. Le codage chromatique est une idée relativement récente dans « l'ingénierie des facteurs humains », mais ses vertus sont à présent largement reconnues. Les hôpitaux tracent sur le sol des corridors des lignes colorées, simplifiant ainsi les directions que doivent suivre les patients : « pour aller à la physiothérapie, suivez la ligne jaune ; pour aller à la transfusion sanguine, suivez la ligne rouge ! » Les fabricants de téléviseurs, d'ordinateurs et autres composants électroniques procèdent à un codage coloré des vastes réseaux de fils qui se trouvent à l'intérieur des postes, qui leur permet de suivre ces fils facilement à la trace. Ce sont là des applications récentes, mais l'idée est évidemment bien plus ancienne ; plus ancienne que la Lettre Écarlate qui servait à marquer l'époux adultère, plus ancienne que les uniformes de

couleur qui servaient à distinguer l'ami de l'ennemi dans le feu de la bataille, plus anciennes en fait, que l'espèce humaine.

Nous avons tendance à penser que le codage chromatique est une manière astucieuse d'introduire aux schèmes de couleur « conventionnels » conçus pour tirer avantage de la vision chromatique « naturelle », mais c'est oublier le fait que la vision chromatique « naturelle » a co-évolué, *d'emblée* ; avec des couleurs dont la raison d'être était le codage chromatique (Humphrey, 1976). Certaines choses dans la nature « avaient besoin d'être vues » et d'autres avaient besoin de les voir ; un système a donc évolué qui avait tendance à minimiser la tâche du dernier en insistant sur l'importance du premier. Considérons les insectes. Leur vision chromatique a co-évolué avec les couleurs des plantes qu'ils ont pollinisées, un bon truc de conception qui a profité aux deux. Sans le codage chromatique des fleurs, la vision des couleurs des insectes n'aurait pas évolué, et vice versa. Le principe de codage chromatique est donc la base de la vision des couleurs chez les insectes, et pas uniquement une invention récente d'une espèce intelligente de mammifères. On pourrait raconter des histoires semblables sur l'évolution de la vision des couleurs dans d'autres espèces. Alors qu'une sorte de vision des couleurs peut avoir évolué initialement dans l'exercice visuel de discrimination des phénomènes inorganiques, on ne sait pas encore clairement si cela s'est produit avec la moindre espèce sur cette planète. (Evan Thompson m'a fait remarquer que les abeilles à miel peuvent utiliser leur espèce particulière de vision chromatique en navigation pour discriminer la lumière solaire polarisée par temps nuageux, mais s'agit-il d'une utilisation secondaire de la vision chromatique qui a originellement co-évolué avec les couleurs des fleurs ?)

Différents systèmes de vision chromatique ont évolué indépendamment, parfois avec des espaces chromatiques radicalement différents. (Pour un bref aperçu, et pour les références, voir Thompson, Palacios et Varela, à paraître.) Les êtres qui ont des yeux ne sont pas tous dotés d'un système ou un autre de vision chromatique. Les oiseaux, les poissons, les reptiles et les insectes ont à l'évidence une vision chromatique, qui ressemble plutôt à notre système « trichromatique » (rouge-vert-bleu) ; les chiens et les chats ne l'ont pas. Parmi les mammifères, seuls les primates ont une vision chromatique, et il y a entre eux d'énormes différences. Quelles sont les espèces qui ont une vision chromatique

et pourquoi ? Cela finit par devenir une histoire fascinante et complexe, et pour une bonne part encore, spéculative.

Pourquoi les pommes deviennent-elles rouges en mûrissant ? Il est naturel de supposer que la réponse se trouvera dans les changements chimiques qui se produisent quand le sucre et d'autres substances atteignent des concentrations variées dans le fruit en train de mûrir, causant diverses réactions, et ainsi de suite. Mais c'est laisser dans l'ombre le fait qu'il n'y aurait pas de pommes pour commencer s'il n'y avait pas de semeurs de graines-mangeurs de pommes pour les voir : aussi le fait que les pommes soient aisément visibles à au moins certaines variétés de mangeurs de pommes est-il une condition de leur existence, pas un simple « hasard » (du point de vue de la pomme !). Le fait que les pommes aient les propriétés de réflectance spectrale de surface qu'elles ont est autant fonction des photopigments qui étaient susceptibles d'équiper les cellules du cône oculaire des fructivores que l'effet des interactions entre le sucre et d'autres composés chimiques du fruit. Les fruits qui n'ont pas de code chromatique sont en piteuse compétition sur les étagères du supermarché de la nature, mais toute publicité mensongère sera punie ; les fruits qui sont mûrs (très nutritifs) *et qui publicisent ce fait* se vendront mieux, mais la publicité doit être taillée à la mesure des capacités et inclinations visuelles des consommateurs que l'on a ciblés.

Au début, les couleurs étaient faites pour être vues de ceux qui étaient faits pour les voir. Mais cela a progressivement évolué, au gré des circonstances, au fur et à mesure où l'on tirait brillamment avantage de tous les matériaux qui pouvaient se trouver sous la main, ce qui explosait à l'occasion en une profusion de trucs nouveaux que l'on inventait, mais avec toujours une tolérance pour bon nombre de variations inutiles *et* de phénomènes constants inutiles (de pure coïncidence). Ces phénomènes constants de coïncidence ont souvent concerné des traits « plus fondamentaux » du monde physique. Une fois qu'il y avait des êtres qui pouvaient distinguer des baies rouges de baies vertes, ils pouvaient aussi distinguer des rubis rouges d'émeraudes vertes, mais c'était juste un bonus de pure coïncidence. Le fait qu'il y ait une différence *de couleur* entre les rubis et les émeraudes peut ainsi être considéré comme un phénomène de couleur *dérivé*. Pourquoi le ciel est-il bleu ? Parce que les pommes sont rouges et les raisins pourpres, et non l'inverse.

C'est une erreur de penser qu'il y a d'abord eu des couleurs – des rochers colorés, de l'eau colorée, un ciel coloré, de la rouille orange rougeâtre et du bleu cobalt clair – et que Dame Nature est arrivée et a tiré parti de ces propriétés en les utilisant comme un codage chromatique des choses. C'est plutôt qu'il y a eu d'abord diverses propriétés de réflectance de surfaces, des propriétés réactives de photopigments, et ainsi de suite, et que Dame Nature a développé à partir de ce matériau brut des systèmes de vision « chromatique »/ de codage « chromatique » ajustés l'un à l'autre, et parmi les propriétés qui sont sorties de cette organisation naturelle générale, on trouve les propriétés que nous autres êtres humains normaux appelons les couleurs. S'il se trouve que le bleu du cobalt et le bleu d'une aile de papillon vont bien ensemble (chez des êtres humains normaux), ce n'est que pure coïncidence, un effet secondaire négligeable des processus qui ont fait naître la vision chromatique et qui par là (comme Locke lui-même aurait pu le reconnaître) ont consacré un certain ensemble curieusement trafiqué de toute une série de propriétés premières avec la propriété secondaire partagée de produire un effet commun sur un ensemble d'observateurs normaux.

« N'empêche, m'objecterez-vous, bien avant qu'il y ait des animaux doués de vision chromatique, il y avait de glorieux couchers de soleil rouges, et de lumineuses émeraudes vertes ! » Eh bien soit ! Mais alors ces mêmes couchers de soleil étaient aussi éblouissants, multicolores, et dégoûtants, s'exprimant en des couleurs que nous ne pouvons voir, et pour lesquels nous n'avons donc pas de noms. Bref, c'est ce qu'il vous faut admettre, s'il y avait ou *pouvait* y avoir des êtres sur une planète dont l'appareil sensoriel se trouvait ainsi affecté par eux. Et pour autant que nous le sachions, il existe quelque part des espèces qui voient naturellement qu'il y a deux (ou dix-sept) couleurs différentes dans une série d'émeraudes que *nous* nous avons trouvées indistinctement vertes.

Bien des êtres humains sont aveugles aux couleurs rouge et verte. Supposons que nous le soyons tous ; il serait alors commun de reconnaître que les rubis et les émeraudes sont tous deux « verouge » (*gred*) – après tout, ils ne paraissent à l'observateur normal que comme d'autres choses rouvertes : des voitures de pompiers, des pelouses bien arrosées, des pommes mûres et non mûres (Dennett, 1969). Si des gens comme nous se mettaient à soutenir que les rubis et les émeraudes sont en fait de couleur

différente, il n'y aurait aucun moyen de déclarer que l'un de ces systèmes de vision des couleurs est « plus vrai » que l'autre.

Le philosophe Jonathan Bennett (1965) attire notre attention sur un cas qui fait ressortir le même point, de façon plus convaincante, sur une autre modalité sensorielle. La substance phénolthiurée, nous dit-il, a un goût plus amer pour un quart de la population humaine et n'a aucun goût pour le reste. Le genre de goût qu'elle a est génétiquement déterminé. Est-ce que la phénolthiurée est plus amère ou sans goût ? Par « eugénie » (contrôle de la race) ou ingénierie génétique, nous pourrions réussir à éliminer le génotype qui fait trouver plus amer le phénol. Si nous y parvenions, la phénolthiurée serait alors *paradigmatiquement* sans goût, aussi dépourvue de goût que de l'eau distillée : sans goût pour tous les êtres humains normaux. Si nous effectuions l'expérience génétique opposée, nous pourrions alors rendre la phénolthiurée *paradigmatiquement* amère. Or, avant l'apparition des êtres humains, la phénolthiurée était-elle *à la fois* amère et sans goût ? Elle était chimiquement semblable à ce qu'elle est maintenant.

Les faits relatifs aux qualités secondes sont inévitablement liés à une classe d'observateurs de référence, mais il y a des manières faibles et fortes de traiter le lien. Nous pouvons dire que les qualités secondes sont *charmantes* plutôt que *suspectes*. Une personne pourrait être charmante, sans qu'aucun observateur susceptible de la trouver charmante ne l'ait encore, en la circonstance, observée, mais elle ne pourrait pas être suspecte – c'est une question de logique – tant que quelqu'un ne l'aurait pas de fait suspectée de quelque chose. On peut dire que des cas particuliers de qualités charmantes (telle que la qualité de charme) existent à titre de dispositions lockéennes avant le moment (s'il y en a un) où elles exercent leur pouvoir sur un observateur, y produisant l'effet défini. Ainsi une femme qui n'aurait jamais été vue (parce qu'elle aurait, supposons, grandi toute seule sur une île déserte) pourrait être authentiquement charmante, dotée du pouvoir dispositionnel d'affecter les observateurs normaux d'une certaine classe de telle ou telle manière, bien qu'elle n'ait jamais eu l'occasion de le faire. Mais on ne peut définir les qualités charmantes indépendamment des inclinations, des susceptibilités, ou des dispositions d'une classe d'observateurs ; cela n'a donc aucun sens de parler de l'existence de propriétés charmantes tout à fait indépendamment de l'existence d'observateurs pertinents.

En vérité, c'est un peu trop fort. Les qualités charmantes ne *seraient* pas définies, cela n'aurait aucun sens de *les* définir – à l'inverse de toutes les autres propriétés trafiquées logiquement possibles – indépendamment d'une telle classe d'observateurs. Ainsi, bien qu'il soit logiquement possible (« rétrospectivement » pourrait-on dire) de rassembler des cas de propriétés chromatiques par quelque chose comme l'énumération pure et dure, les raisons de sélectionner de telles propriétés (pour expliquer, par exemple, certaines régularités causales dans un ensemble d'objets curieusement compliqués) dépendent de l'existence de la classe des observateurs.

Est-ce que les éléphants de mer sont charmants ? Pas pour nous. Il est difficile d'imaginer un être plus laid. Ce qui rend un éléphant de mer charmant pour un autre éléphant de mer n'est pas ce qui rend une femme charmante pour un homme, et appeler charmante une femme non encore observée qui, en la circonstance, pourrait exercer une forte attirance sur des éléphants de mer serait commettre un abus et sur elle et sur les termes. Ce n'est que par référence aux goûts humains qui sont des caractères contingents et en vérité idiosyncrasiques du monde, que l'on peut identifier la propriété de charme (pour-un-être-humain).

D'autre part, des qualités suspectes (telles que la propriété d'être suspect) se comprennent de telle manière qu'elles présupposent que toute occurrence de la propriété a déjà eu son effet définitionnel sur au moins un observateur. Vous pouvez être éminemment digne de soupçon – vous pouvez même être manifestement coupable – mais vous ne pouvez être suspect tant que quelqu'un ne vous soupçonne pas effectivement. Je ne prétends pas que les couleurs soient des qualités suspectes. Il ne s'agit pas de nier notre intuition que l'émeraude non encore observée au milieu d'un bloc de minerai est *déjà* verte. Mais je soutiens que les couleurs sont des qualités charmantes dont l'existence, tributaire comme elle l'est de la classe de référence des observateurs, n'a aucun sens dans un monde où les observateurs n'ont pas leur place. C'est plus facile à accepter pour certaines qualités secondes que pour d'autres. Que les vapeurs sulfureuses s'exhalant des volcans primitifs aient été jaunes paraît relativement plus objectif que le fait qu'elles sentent mauvais, mais tant que ce que nous voulons dire par « jaune » est ce que *nous* voulons dire par « jaune », les deux affirmations sont parallèles. Car supposons qu'un séisme pri-

mitif ait soulevé le versant d'une colline mettant à nu les strates formées de centaines de couches chimiquement différentes. Ces strates étaient-elles visibles ? Nous devons nous demander à qui. Peut-être que certaines d'entre elles seraient visibles pour nous, et d'autres pas. Peut-être que certaines des strates invisibles seraient visibles à des pigeons tétrachromatiques, ou à des êtres qui ont vu la partie infrarouge ou ultraviolette du spectre électromagnétique. Pour la même raison, on ne peut de façon sensée se demander si la différence entre les émeraudes et les rubis est visible sans spécifier le système visuel concerné.

L'évolution amortit le « subjectivisme » ou le « relativisme » implicite dans le fait que les qualités secondes soient des qualités charmantes. Elle montre que l'absence de caractères communs « simples » ou « fondamentaux » dans des choses qui sont toutes de la même couleur n'est pas un signe distinctif de totale illusion, mais plutôt un signe de tolérance répandue à l'égard des révélations « fausses positives » sur les propriétés écologiques qui comptent vraiment ². Les catégories fondamentales de notre espace chromatique (et bien sûr de notre espace olfactif et de notre espace sonore, etc.) sont le produit des pressions sélectives, si bien qu'en général cela a un sens de demander à quoi sert une discrimination ou préférence donnée. Il y a des raisons pour lesquelles nous évitons les odeurs de certaines choses et en recherchons d'autres, pour lesquelles nous préférons certaines couleurs à d'autres, pour lesquelles certains sons nous ennuiant plus que d'autres, ou nous apaisent davantage. Il se peut que ce ne soient pas toujours *nos* raisons, mais plutôt les raisons de lointains ancêtres, laissant leurs traces fossilisées dans les prédispositions constitutives qui façonnent de façon innée nos espaces qualitatifs. Mais en bons darwiniens, nous devrions aussi reconnaître la possibilité – en vérité, la nécessité – d'autres prédispositions, non fonctionnelles, distribuées fortuitement à travers la population selon des variations génétiques. Pour que la pression sélective favorise de façon différenciée ceux qui manifestent une prédisposition à l'encontre de F une fois que F devient écologiquement important, c'est qu'il doit y avoir eu une variation inutile (non encore fonctionnelle) dans « l'attitude envers F » sur laquelle peut agir la sélection. Par exemple, si le fait de manger des tripes *devait* impliquer un destin néfaste préreproductif dans le futur, seuls ceux d'entre nous qui sont « naturellement » (et *désormais*

inutilement) disposés contre le fait de manger des tripes auraient un avantage (peut-être léger au départ, mais rapidement susceptible de devenir explosif, si les conditions le favorisaient). Ce n'est donc pas parce que vous trouvez indescriptiblement et ineffablement abominable quelque chose (par exemple les brocolis), qu'il y a une raison à cela. Pas plus que vous n'êtes déficient si vous n'êtes pas d'accord là-dessus avec vos semblables. Il se peut qu'il s'agisse tout simplement de l'une des propensions innées de votre espace qualitatif qui n'a, pour le moment, aucune espèce de signification fonctionnelle. (Et dans votre intérêt, vous feriez mieux d'espérer que s'il doit jamais y en avoir une, c'est parce que les brocolis se sont brusquement révélés nocifs pour nous.)

Ces considérations évolutionnistes aident considérablement à comprendre pourquoi les qualités secondes se révèlent être aussi « ineffables », aussi réfractaires à toute définition. Comme la propriété formelle M du morceau de carton des Rosenberg, les qualités secondes sont extrêmement réfractaires à une définition directe. Il est de l'essence du tour des Rosenberg que nous ne puissions remplacer notre prédicat postiche « M » par une description plus longue, plus complexe, mais exacte et exhaustive de la propriété, car si nous le pouvions, nous (ou quelqu'un d'autre) pourrions utiliser cette description comme une recette pour produire une autre instance de M ou d'un autre détecteur de M. Nos détecteurs de qualités secondes n'étaient pas spécifiquement chargés de ne détecter que des propriétés difficiles à définir, mais le résultat revient quasiment au même. Comme le fait observer Akins (1989), nos systèmes sensoriels *ne sont pas faits pour* détecter des propriétés « fondamentales » ou « naturelles » de l'environnement, mais simplement pour servir notre but « narcissique » de rester en vie ; la nature ne construit pas de moteurs épistémiques.

La seule manière *qu'on ait sous la main* de dire quelle propriété formelle est M c'est d'indiquer le détecteur de M et de dire que M est la propriété formelle que cette chose-là détecte. La même difficulté se présente naturellement à quiconque essaie de dire quelle propriété quelqu'un détecte (ou détecte par erreur) lorsque quelque chose « lui apparaît telle qu'elle lui apparaît ». On peut donc à présent répondre à la question que nous posions au début de ce paragraphe : quelle propriété Otto juge-t-il qu'a une chose lorsqu'il juge qu'elle est rose ? La propriété qu'il appelle rose.

Et de quelle propriété s'agit-il ? C'est difficile à dire, mais cela ne devrait pas nous embarrasser, parce que nous pouvons dire pourquoi c'est difficile. Le mieux que nous puissions faire, lorsqu'on nous demande quelles propriétés de surface nous détectons par vision chromatique, c'est de dire, sur un mode qui n'est pas très informatif, que nous détectons les propriétés que nous détectons. Si quelqu'un veut quelque chose de plus informatif sur ces propriétés, il existe une littérature immense et plutôt incompressible en biologie, en neurosciences et en psychophysique à consulter. Quant à Otto, il ne peut rien dire de plus sur la propriété qu'il appelle rose en disant « C'est *ceci* ! » (montrant par là qu'il désigne comme étant « à l'intérieur » une propriété phénoménale de son expérience). Tout ce qu'accomplit (au mieux) ce mouvement, c'est d'indiquer à son propre état idiosyncrasique de discrimination chromatique, un mouvement qui est parallèle à celui qui consiste à brandir un morceau de carton et à dire qu'il détecte *cette* propriété formelle. Otto indique peut-être bien son procédé de discrimination, mais pas un quelconque *quale* que celui-ci ferait transparaître, ou prendrait à son compte, ou exprimerait, quand il fait son travail. Il n'existe rien de tel.

– Tout de même ! [insiste Otto]. Vous n'avez toujours pas dit pourquoi il faudrait que le rose apparaisse comme *ceci* !

– Comme quoi ?

– Comme ceci. Comme le rose particulièrement ineffable, merveilleux, intrinsèque que je suis en train actuellement d'apprécier. Ce n'est pas *là* une propriété de réflectance de surface des objets extérieurs pleine de circonvolutions indescriptibles.

– Je vois, Otto, que tu utilises le terme « apprécier ». Tu n'es pas le seul. Souvent, quand un auteur veut insister sur le fait qu'on est passé d'un argument de (pure) neuro-anatomie à l'expérience, de la (pure) psychophysique à la conscience, de la (pure) information aux qualia, on fait entrer en scène le mot « apprécier ».

Apprécier nos expériences

« Mais Dan, les *qualia*, c'est ce qui fait que la vie vaut la peine d'être vécue ! »

Wilfrid Sellars (autour d'une bonne bouteille de Chambertin), Cincinnati, 1971.

« Si ce que je veux quand je bois du bon vin c'est de l'information sur ses propriétés chimiques, pourquoi est-ce que je ne me contente pas de lire l'étiquette ? »

Sydney Shoemaker, Colloque de Tufts, 1988.

Certaines couleurs sont faites pour être aimées, et aussi certaines odeurs et certains goûts. Et d'autres couleurs, odeurs et goûts sont faits pour ne pas l'être. Pour dire les choses en termes plus précis, ce n'est pas un hasard si nous (et d'autres êtres capables de les détecter) aimons ou n'aimons pas les couleurs, les odeurs, les goûts et autres qualités secondes. Tout comme nous avons hérité de détecteurs évolués de symétrie verticale dans le système visuel qui est propre à nous avertir (comme nos ancêtres) du fait écologiquement significatif qu'un autre être est en train de nous regarder, nous avons de même hérité de détecteurs de qualités évolués qui ne sont pas des reporters désintéressés, mais plutôt des signalisateurs et des sentinelles, des sirènes au sens des pompiers et au sens homérique du terme.

Comme nous l'avons vu au chapitre 6, consacré à l'évolution, ces signaux d'alarme originaires ont par la suite été cooptés en une foule d'organisations plus compliquées, construites à partir de millions d'associations, et formées, dans le cas des humains, par des milliers de mêmes. De cette manière, l'appel brutal du style « va vite le chercher » du sexe et de la nourriture, comme l'aversion brutale, du style « à fuir de toute urgence » de la douleur et de la peur, se trouvent fondus en un mélange composé de toutes sortes de combinaisons piquantes. Quand un organisme découvre que cela est payant d'être attentif à certaines composantes du monde *en dépit* de son aversion constitutive à le faire, il lui faut construire une contre-attaque quelconque pour empê-

cher l'aversion de gagner la partie. La tension à moitié stable qui en résulte peut alors elle-même devenir un goût acquis, qu'il convient de rechercher dans certaines conditions. Quand un organisme découvre qu'il doit étouffer les effets de certains signaux insistants s'il veut suivre la conduite appropriée, il lui faut cultiver un goût pour toutes les séquences d'activités possibles qui soient susceptibles de produire la paix et le repos désirés. C'est ainsi que nous pourrions nous mettre à adorer la nourriture épicée qui met la bouche en feu (Rozin, 1982), la musique délicieusement « discordante », et simultanément, le réalisme calme et rafraîchissant d'Andrew Wyeth et l'expressionnisme brûlant et déstabilisant de Willem de Kooning. Marshall McLuhan (1967) a proclamé que le véhicule est le message, quasi-vérité qui peut être plus vraie dans le système nerveux que dans n'importe quel autre forum de communication. Ce que nous voulons quand nous sirotons un grand vin, ce n'est pas, en vérité, de l'information concernant son contenu chimique ; ce que nous voulons, c'est *être informés* sur son contenu chimique, en suivant notre manière favorite. Et notre préférence s'appuie *en dernière analyse* sur les inclinations qui sont toujours câblées dans notre système nerveux, même si leur signification écologique peut avoir disparu il y a des années-lumière.

Notre propre technologie nous a, pour une bonne part, caché ce fait. Comme le note le psychologue Nicholas Humphrey :

« Quand je regarde tout autour de la pièce où je travaille, d'innombrables couleurs qui sont l'œuvre de l'homme me sautent aux yeux en provenance de toutes les surfaces : des livres, des coussins, le tapis sur le plancher, une tasse, une boîte d'agrafes – bleu clair, rouges, jaunes, vertes. Il y a autant de couleur, ici, que dans n'importe quelle forêt tropicale. Pourtant, alors que presque chaque couleur aurait un sens dans la forêt, ici dans mon bureau, presque rien n'en a. L'anarchie chromatique a pris le dessus » (1983, p. 149).

Considérons par exemple le fait curieux que les singes n'aiment pas la lumière rouge. Si on leur en donne le choix, les singes rhésus montrent une forte préférence pour l'extrémité bleu-vert du spectre, et deviennent agités lorsqu'ils doivent endurer de longs moments dans des environnements rouges (Humphrey, 1972, 1973, 1983 ; Humphrey et Keeble, 1978). Pourquoi ? Humphrey fait remarquer que le rouge sert toujours comme signal d'alarme, que c'est la couleur extrême du codage chromatique, mais que, pour cette raison même, il est ambigu : le fruit rouge

peut être bon à manger, mais le serpent ou l'insecte rouge est probablement en train de signaler qu'il est venimeux. « Rouge » envoie donc des messages mixtes. Mais pourquoi envoie-t-il, pour commencer, un message d'« alarme » ? Peut-être parce qu'il s'agit du plus fort contraste dont on dispose avec l'arrière-plan ambiant de végétation verte ou de mer bleue, ou – dans le cas des singes – parce que la lumière rouge (de rouge à rougeâtre – d'orange à orange clair) est la lumière du crépuscule et de l'aube, moments du jour où pratiquement tous les prédateurs de singes se livrent à la chasse.

Les propriétés affectives ou émotionnelles du rouge ne se limitent pas aux singes rhésus. Tous les primates partagent ces réactions, y compris les êtres humains. Si vos travailleurs d'usine s'attardent trop longtemps dans les toilettes, peindre les murs des toilettes en rouge résoudra ce problème – mais en créera d'autres. (Voir Humphrey, 1992.) De telles réponses « viscérales » ne se limitent évidemment pas aux couleurs. La plupart des primates élevés en captivité et qui n'ont jamais vu de serpent manifesteront immanquablement que les serpents leur répugnent au moment où ils en voient un, et il est probable que le dégoût traditionnel des hommes pour les serpents a une source biologique qui explique la source biblique, plutôt que l'inverse³. En un mot, notre héritage génétique vient tout déranger et favoriser les mêmes de dégoût pour les serpents.

Voici deux explications différentes du sentiment de malaise que la plupart d'entre nous éprouvent (même si nous l'« attrapons ») à la vue d'un serpent :

1. Les serpents évoquent en nous un *quale* particulier intrinsèque d'envie de vomir quand nous les regardons, et notre sentiment de malaise est une réaction à ce *quale*.

2. Nous nous trouvons moins enclins à voir des serpents en raison de prédispositions innées constitutives de notre système nerveux. Celles-ci favorisent la décharge d'adrénaline, font intervenir des routines de fuite, et, en activant diverses liaisons associatives, font entrer en jeu une foule de scénarios impliquant le danger, la violence et le dommage. L'aversion originaire de primate est en nous transformée, révisée, déviée d'une centaine de manières par les mêmes qui l'ont exploitée, cooptée, façonnée. (Nous pourrions formuler une explication de ce type « fonctionnaliste » à bien des niveaux différents. Par

exemple, nous pourrions nous laisser aller à parler de façon plus libérale du pouvoir des perceptions de serpents à produire angoisses, craintes, anticipations de douleur, et ainsi de suite, mais on pourrait penser qu'on « triche » en procédant ainsi, aussi l'éviterai-je.)

Le problème avec le premier genre d'explication c'est qu'elle a seulement l'air d'être une explication. L'idée qu'une propriété « intrinsèque » (de rose occurrent, d'envie de vomir devant les serpents, de douleur, d'arôme de café) pourrait *expliquer* les réactions d'un sujet est désespérante — c'est un cas manifeste de *virtus dormitiva* (voir page 86, chapitre 2, quatrième section). Mais accuser une théorie d'abriter une *virtus dormitiva* n'est pas si simple. Parfois, il est parfaitement bien venu de postuler une *virtus dormitiva* temporaire, en attendant plus ample investigation. La conception est, par définition, pourrait-on dire, la cause de la grossesse. Si nous n'avions aucun autre moyen d'identifier la conception, dire à quelqu'un qu'elle est devenue enceinte parce qu'elle a conçu serait un geste vide, non une explication. Mais une fois que nous avons réussi à nous faire une idée de la théorie mécanique requise pour qu'il y ait conception, nous voyons *comment* la conception est cause de la grossesse, et l'information passe à nouveau. Dans le même esprit, nous pourrions identifier les qualia, par définition, comme les causes prochaines de nos joies et de nos souffrances (pour le dire grossièrement), puis nous mettre à nous décharger de notre obligation d'information en poursuivant le second style d'explication. Mais assez curieusement, les qualophiles (comme j'appelle ceux qui croient encore aux qualia) ne veulent rien savoir ; ils soutiennent, comme Otto, que les qualia « réduits » à de simples complexes de dispositions mécaniquement accomplies à réagir ne sont pas les qualia dont ils parlent. *Leurs* qualia sont quelque chose de différent.

Considérons [dit Otto] la manière dont l'anneau rose m'apparaît à moi, juste *en ce moment*, en ce moment précis, séparé de toutes mes dispositions, des associations passées et des activités futures. C'est *cela*, la *manière* purifiée, séparée *dont il se présente à moi* relativement aux couleurs en ce moment — c'est cela mon quale rose.

Otto a simplement commis une erreur. En fait, c'est là la grosse erreur, la source de tous les paradoxes sur les qualia,

comme nous le verrons. Mais avant de montrer pourquoi c'est de la folie de suivre cette voie, je voudrais faire la démonstration de certains des bénéfices positifs que présente la voie que rejette Otto : la voie « réductionniste » qui consiste à *identifier* « la manière dont cela se présente à moi » avec la somme totale de toutes les dispositions réactives idiosyncrasiques inhérentes à mon système nerveux qui résultent de ma confrontation avec un certain schéma de stimulation.

Considérons quel effet cela a dû faire à un pratiquant luthérien de Leipzig en, disons, 1725, d'entendre l'une des cantates chorales de J. S. Bach la première fois qu'elle a été jouée. (Cet exercice consistant à imaginer *quel effet cela peut faire* est une manière de s'échauffer pour le chapitre 13, où nous traiterons de la conscience chez d'autres animaux.) Il n'y a probablement aucune différence biologique significative entre nous aujourd'hui et les luthériens allemands du XVIII^e siècle ; nous sommes de la même espèce, et fort peu de temps s'est écoulé. Mais, en raison de l'énorme influence de la culture — la mémosphère —, notre univers psychologique est tout à fait différent du leur, d'une manière qui aurait un impact notable sur nos expériences respectives à la première écoute d'une cantate de Bach. Notre imagination musicale s'est enrichie et compliquée de bien des manières (grâce à Mozart, Charlie Parker, les Beatles), mais elle a aussi perdu certaines des puissantes associations sur lesquelles Bach pouvait compter. Ses cantates chorales étaient construites autour des chorals, hymnes mélodiques traditionnels qui étaient profondément familiers aux pratiquants de son époque, et qui provoquaient des vagues d'association émotionnelle et thématique dès qu'on pouvait en voir apparaître les traces ou les échos dans la musique. La plupart d'entre nous aujourd'hui ne connaissent ces chorals que par les arrangements qu'en a faits Bach, si bien que lorsque nous les entendons nous les entendons d'une oreille différente. Si nous voulons imaginer quel effet cela pouvait faire d'être un auditeur de Bach à Leipzig, il ne nous suffit pas d'entendre les mêmes tons sur les mêmes instruments et dans le même ordre. Il nous faut aussi nous préparer en quelque sorte à répondre à ces tons avec les mêmes battements de cœur, les mêmes frémissements, et les mêmes ondes nostalgiques.

Il n'est pas complètement impossible de nous préparer ainsi. Un musicologue qui aurait soigneusement évité tout contact avec la musique postérieure à 1725 et qui se serait profondément

familiarisé avec la musique traditionnelle de cette période en serait une bonne première approximation. Plus important encore, comme le montrent ces observations, il n'est pas impossible de savoir exactement de quelle manière nous aurions à nous préparer, que nous ayons ou non l'intention de nous donner toute cette peine. Nous pourrions donc savoir quel effet cela peut faire « dans l'abstrait » pour ainsi dire, et en fait je viens de vous le dire : les habitants de Leipzig, entendant les cantates chorales, étaient portés à se souvenir des associations qui avaient déjà donné une saveur particulière à leur reconnaissance de mélodies chorales. Il est assez facile d'imaginer quel effet *cela* a dû leur faire — aux variations près dues à notre expérience propre. Nous pouvons imaginer quel effet cela ferait d'entendre l'arrangement de Bach de chants de Noël familiers, par exemple, ou bien de « Ma cabane au Canada ». Nous ne pouvons pas dans le détail nous le représenter, mais c'est seulement parce que nous ne pouvons oublier ou laisser de côté tout ce que nous savons et que ne savaient pas les habitants de Leipzig.

Pour voir à quel point cet excédent de bagage que nous avons est crucial, imaginons que des musicologues dénichent une cantate de Bach jusqu'alors inconnue, qui est bien du grand homme, sans nul doute aucun, mais qui se trouvait dans le fond d'un tiroir et qui n'avait probablement jamais encore été entendue, pas même par le compositeur lui-même. Tout le monde brûlerait de l'entendre, de faire l'expérience, pour la première fois, des « qualia » que les habitants de Leipzig auraient éprouvés, si seulement ils l'avaient entendue, mais voilà que cela se révèle impossible, parce que le thème principal de la cantate, par une horrible coïncidence, se trouve être les sept premières notes de « Debout les gars, réveillez-vous ! ». Nous qui avons les oreilles pleines de cet air ne pourrions *jamais* entendre la version de Bach comme il l'aurait voulu, ou comme les habitants de Leipzig l'auraient reçue.

On pourrait difficilement trouver un cas plus clair de blocage de l'imagination, mais on notera que cela n'a rien à voir avec des différences biologiques ou même avec des propriétés « intrinsèques » ou « ineffables » de la musique de Bach. La raison pour laquelle nous ne pourrions revivre en imagination et de façon détaillée (et précise) l'expérience musicale des habitants de Leipzig est simplement que nous aurions à nous embarquer pour le voyage imaginaire, et que nous en savons trop. Mais si nous le

voulons, nous pouvons soigneusement dresser la liste des différences qui existent entre nos dispositions et connaissances et les leurs, et en comparant les listes, nous mettre à apprécier, *avec tous les détails voulus*, les différences qu'il y a entre l'effet que cela pouvait leur faire d'écouter du Bach, et l'effet que cela peut nous faire à nous. Même si nous pouvons nous lamenter sur le caractère inaccessible de ces expériences, nous pouvons du moins le comprendre. Il ne resterait pas de mystère ; juste une expérience que l'on pourrait décrire de façon très exacte, mais qu'on ne pourrait apprécier directement à moins de nous livrer à un travail ridicule de reconstruction de nos structures dispositionnelles personnelles.

Mais les qualophiles n'ont pas voulu de cette conclusion. Il leur a semblé que même si une investigation comme celle que nous venons d'imaginer pouvait résoudre *presque* toutes les questions que nous nous posions sur l'effet que cela peut faire d'être des habitants de Leipzig, il y aurait un résidu ineffable, *quelque chose* quant à l'effet que cela fait à des habitants de Leipzig qu'aucun progrès encore envisageable sur la connaissance « dispositionnelle » et « mécaniste » ne pourrait réduire à zéro. C'est pourquoi les qualophiles doivent prendre les qualia comme des caractères *en plus*, qui sont au-delà et strictement indépendants du câblage qui détermine le fait de battre en retraite, de froncer les sourcils, de hurler, et autres « simples comportements » de dégoût, de répulsion, de peur. Cela ressort clairement si nous revenons à notre exemple des couleurs.

Supposons que nous suggérions à Otto que ce qui a fait que son « rose occurrent » soit l'expérience provocante particulière qu'il a appréciée n'était que la résultante de toutes les associations innées et acquises et des dispositions réactives déclenchées par la manière particulière dont il a été informé (ou mal informé) par ses yeux :

Ce que *sont* les qualia, Otto, ce sont tout simplement ces complexes de dispositions. Quand vous dites : « Ceci est mon quale », ce que vous sélectionnez, ou désignez, *que vous vous en rendiez compte ou non*, c'est votre complexe idiosyncrasique de dispositions. Vous *avez l'impression* de désigner une chose ou une autre de privée, d'ineffable dans l'œil de votre esprit, une nuance privée de rose homogène, mais c'est seulement une impression, et non la réalité. Votre « quale » est un personnage qui a bonne allure dans le monde fictionnel de votre hétérophénoménologie,

mais ce qu'il se révèle être dans le monde *réel* de votre cerveau n'est qu'un complexe de dispositions.

Il n'est pas possible que ce soit là le fin mot de l'histoire [rétorque Otto, emboîtant fatalement le pas à la tradition qualophile], car même si ce complexe de simples dispositions était peut-être la base ou la source, ou autre, de mon *quale* particulier de rose, ils pourraient tous être changés sans changer mon *quale* intrinsèque, ou mon *quale* intrinsèque pourrait changer sans changer cette multitude de simples dispositions. Par exemple, mes *qualia* pourraient être *inversés* sans inverser toutes mes dispositions. Je pourrais avoir toutes les réactivités et associations que j'ai à présent pour le *vert* en accompagnement du *quale* que j'ai à présent pour le *rouge*, et vice versa.

Une fantaisie philosophique : les *qualia* inversés

L'idée de la possibilité de tels « *qualia* inversés » est l'un des mêmes les plus virulents de la philosophie. Locke l'a discutée dans son *Essai sur l'entendement humain* (1690) et bon nombre de mes étudiants me disent que lorsqu'ils étaient enfants, ils sont tombés eux-mêmes tout seuls sur la même idée, qui les a fascinés. L'idée paraît totalement claire et inoffensive :

Il y a les manières dont les choses m'apparaissent à l'œil, à l'oreille, au nez, et ainsi de suite. Voilà qui est évident. Mais je me demande si la manière dont les choses m'apparaissent est la même que celle dont les choses apparaissent à autrui.

Les philosophes ont composé autant de variations diverses sur ce thème, mais la version classique est la version interpersonnelle : comment sais-je que vous et moi voyons la même couleur *subjective* quand nous regardons quelque chose ? Comme notre apprentissage des termes de couleur s'est fait par ostension d'objets colorés publics, notre comportement verbal sera adapté même si nous faisons l'expérience de couleurs subjectives entièrement différentes – même si la manière dont les choses rouges m'apparaissent est la manière dont les choses vertes, par exemple, vous apparaissent. Nous appellerions « rouges » et « vertes » les mêmes choses publiques même si nos expériences privées étaient « tout le contraire » (ou simplement différentes).

Y a-t-il un moyen de dire si tel est le cas ? Considérons l'hy-

pothèse selon laquelle les choses rouges paraissent semblables à vous et à moi. Cette hypothèse est-elle à la fois irréfutable et inconfirmable ? Nombreux sont ceux qui l'ont pensé, et certains en ont conclu que c'est pour cette raison précisément que cela est absurde, en dépit de ce que cela peut avoir d'attirant au départ pour le sens commun. D'autres se sont demandé si la technologie pourrait venir à la rescousse et confirmer (ou infirmer) l'hypothèse de l'inversion interpersonnelle du spectre. Le film de science-fiction *Brainstorms* (qui n'est pas, je m'empresse de le dire, une version de mon livre *Brainstorms*) a exactement mis en scène le dispositif imaginaire idoine : un appareillage neuroscientifique s'adapte à votre tête et alimente votre expérience visuelle dans mon cerveau à travers un câble. Les yeux fermés, je rapporte exactement tout ce que vous regardez, à ceci près que je m'émerveille du jaune du ciel, du rouge de l'herbe, et ainsi de suite. Si nous avions une telle machine, est-ce que le genre d'expérimentation qu'on effectuerait avec elle ne pourrait pas confirmer, empiriquement, l'hypothèse de la différence entre nos qualia ? Mais supposons que le technicien branche le câble de connexion, lui fasse prendre un virage à 180 degrés, le réinsère dans la douille, et que je rapporte maintenant que le ciel est bleu, que l'herbe est verte, et ainsi de suite. Quelle serait la « bonne » orientation de la prise ? Concevoir et construire un tel dispositif – à supposer pour le moment que ce soit possible – exigerait que sa « fidélité » soit accordée et calibrée par la normalisation des rapports des deux sujets, si bien que nous serions ramenés à notre point de départ. On pourrait évidemment éviter cette conclusion en compliquant un peu les choses, mais les qualophiles sont d'accord pour dire que c'est une cause perdue ; il semble y avoir un accord général sur la morale de *cette* expérience de pensée : à savoir qu'aucune comparaison intersubjective de qualia ne serait possible, même avec une technologie parfaite. Toutefois, cela vient étayer la thèse scandaleusement « vérificationniste » ou « positiviste » selon laquelle l'idée même de qualia inversés est absurde – et donc que l'idée même de qualia est absurde. Pour reprendre les termes du philosophe Ludwig Wittgenstein et de sa fameuse analogie du « scarabée dans la boîte » :

« La chose qui est dans la boîte n'a absolument pas de place dans le jeu de langage ; pas même comme un *quelque chose* ; car la boîte pourrait même être vide. – Non, on peut "diviser à travers" par

la chose dans la boîte ; quelle que soit la chose, cela l'annule » (1958, p. 100).

Mais qu'est-ce que cela veut dire au juste ? Est-ce que cela veut dire que les qualia sont réels mais inefficaces ? Ou qu'il n'y a finalement pas de qualia du tout ? Il a toujours semblé évident à la plupart des philosophes qui y ont réfléchi que les qualia sont réels, même si une différence de qualia devait être une différence absolument impossible à entrevoir. Là en étaient difficilement les choses, jusqu'à ce que quelqu'un se mette à rêver de la version censément améliorée de l'expérience de pensée : l'inversion *intrapersonnelle* du spectre. Cette idée semble être venue, indépendamment, à l'esprit de plusieurs personnes (Gert, 1965 ; Putnam 1965 ; Taylor, 1966 ; Shoemaker, 1969 ; Lycan, 1973). Dans cette version, les expériences à comparer se font toutes dans un seul esprit, aussi n'avons-nous pas besoin de la désespérante machine Brainstorm :

Vous vous réveillez un matin et avez l'impression que l'herbe est devenue rouge, le ciel jaune, et ainsi de suite. Personne d'autre ne remarque d'anomalies de couleurs dans le monde ; c'est donc que le problème est en vous. Vous avez le droit, semble-t-il, de conclure que vous avez connu une inversion visuelle chromatique des qualia. Comment cela est-il arrivé ? Pendant que vous dormiez, il se trouve que des neurochirurgiens fous ont inversé tous les fils – les neurones – qui conduisent des cellules du cône sensible à la couleur à vos rétines.

Jusque-là tout va bien. L'effet que cela aurait sur vous serait impressionnant, peut-être même terrifiant. Vous seriez sûrement capables d'entrevoir que la manière dont les choses vous apparaissent à présent est très différente, et nous aurions même une explication scientifique appropriée du phénomène : les groupements neuronaux du cortex visuel qui « se soucient » de la couleur, par exemple, devraient leur stimulation à un ensemble systématiquement inversé de récepteurs rétiniens. La bataille est donc à moitié gagnée, semble-t-il : on pourrait somme toute entrevoir une différence dans les qualia, s'il y avait une différence qui se développait rapidement chez une seule personne⁴. Mais ce n'est que la moitié de la bataille, car le neurochirurgien imaginaire, qui est un farceur, a aussi inversé toutes vos dispositions réactives ; non seulement vous *dites* que vos expériences de couleurs ont toutes été chambardées, mais le comportement non

verbal que vous avez par rapport aux couleurs a été lui aussi inversé. La crispation que vous ne manifestiez qu'à la vue d'une lumière rouge, vous la manifestez à présent à la vue de la lumière verte, et vous avez perdu l'aisance avec laquelle vous vous reposiez chaque jour sur différents schèmes de codage chromatique (si vous jouez au basket-ball pour les Celtes de Boston, vous ne cessez de passer la balle par erreur aux types qui jouent en uniformes rouges).

Ce dont le qualophile a besoin, c'est d'une expérience de pensée qui démontre que la manière dont les choses apparaissent peut être indépendante de toutes ces dispositions réactives. Il nous faut donc compliquer l'histoire en la développant un peu plus avant ; il nous faut décrire quelque chose qui se passe et qui défasse l'inversion des dispositions réactives tout en laissant intacts les « qualia » inversés. C'est ici que toute la littérature qu'on peut trouver là-dessus se lance dans des fantaisies encore plus alambiquées, car personne ne pense un seul instant que la-manière-dont-apparaissent-les-choses soit jamais *réellement* divorcée d'avec les dispositions réactives du sujet ; c'est simplement que les qualophiles jugent qu'il s'agit là d'une importante *possibilité de principe*. Pour le montrer, il leur faut décrire un cas possible, aussi baroque soit-il, dans lequel il serait évident que ce détachement est réel. Considérons une histoire qui ne *marchera pas* :

Une nuit, pendant que vous dormez, des neurochirurgiens fous inversent tous les câbles des cellules du cône (juste comme avant), à la suite de quoi, un peu plus tard, dans la même nuit agitée, arrive une autre équipe de neurochirurgiens, l'équipe B, qui se met à effectuer un recâblage *complémentaire* un peu plus haut sur le nerf optique.

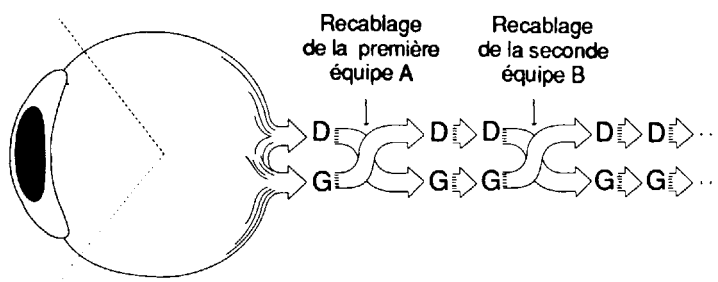


Figure 11.1

Cela rétablit toutes les anciennes dispositions réactives (c'est du moins ce qu'on peut imaginer), mais, hélas, cela rétablit aussi les anciens qualia. Les cellules du cortex qui « se soucient de » la couleur, par exemple, retrouveront à nouveau leurs signaux d'origine, grâce à la réparation rapide du dommage causé par l'équipe B. Le deuxième changement d'aiguillage est survenu trop tôt, semble-t-il ; il est survenu *sur le chemin qui remonte* à l'expérience consciente. Il nous faudra donc raconter l'histoire autrement, le deuxième changement d'aiguillage survenant plus tard, *après* que les qualia inversés ont fait leur entrée dans la conscience, mais *avant* que la moindre des réactions inversées à leur égard ne puisse s'installer. Mais est-ce possible ? Pas si les arguments invoqués en faveur du modèle des Versions Multiples sont corrects. On ne peut tracer aucune ligne à travers la chaîne causale qui va de la pupille à travers la conscience jusqu'au comportement suivant tel que toutes les réactions à x se produisent après elle et que la conscience de x ne se produise pas avant elle. Il en est ainsi parce que ce n'est pas une chaîne causale simple, mais un réseau causal, aux multiples chemins sur lesquels des Versions Multiples ont été éditées simultanément et de manière semi-indépendante. L'histoire du qualophile aurait un sens s'il y avait un Théâtre Cartésien, un endroit bien précis dans le cerveau où se produisait l'expérience consciente. Si un tel lieu existait, nous pourrions le mettre entre crochets avec les deux changements d'aiguillage, en laissant les qualia inversés dans le théâtre, tout en gardant normalisées toutes les dispositions réactives. Mais puisqu'un tel Théâtre Cartésien n'existe pas, l'expérience de pensée n'a aucun sens. Il n'y a aucun moyen cohérent de raconter l'histoire dont on a besoin. Aucun moyen d'isoler les propriétés *présentées* à la conscience à partir des multiples réactions du cerveau à ses discriminations, parce qu'il n'existe aucun autre processus de présentation de ce genre.

Dans les textes qu'on peut lire sur l'inversion du spectre, le deuxième changement d'aiguillage est souvent censé s'accomplir non par opération chirurgicale, mais par une adaptation graduelle du sujet au nouveau régime d'expériences. Cela a apparemment un sens ; les gens peuvent étonnamment bien s'adapter à des modifications bizarres de leur sensibilité. Il y a eu de nombreuses expériences d'inversion du champ visuel dans lesquelles les sujets portent des lunettes qui font voir les choses à l'envers – en redressant complètement l'image rétinienne vers le haut ! (Par exemple Stratton, 1896 ; Kohler, 1961 ; Welch,

1978. Pour un bon résumé, voir aussi Cole, 1990.) Après plusieurs jours où ils n'ont pas quitté tel ou tel genre de lunettes inversantes (cela fait une différence – certaines variétés avaient un large champ de vision, et d'autres donnaient aux voyants une sorte de vision en tunnel), les sujets font preuve d'une capacité d'adaptation époustouflante. Dans le film de Ivo Kohler sur son expérience de Innsbruck, nous voyons deux de ses sujets, qui se trouvent drôlement désarmés quand ils mettent pour la première fois leurs lunettes, descendre des pistes à ski, traverser à vélo des rues embouteillées, tout en continuant à porter les lunettes inversantes auxquelles ils se sont apparemment complètement adaptés.

Supposons donc que vous vous adaptiez progressivement à l'inversion chirurgicale de votre vision chromatique. (La volonté ou la nécessité de s'adapter à cette situation est une tout autre question, mais nous pouvons aussi bien concéder ce point aux qualophiles, pour précipiter leur chute.) À l'évidence, certaines adaptations seraient au départ *post-expérientielles*. Nous pouvons supposer que le ciel clair vous *apparaîtrait* toujours jaune, mais que vous vous mettriez à l'appeler bleu pour marcher au pas avec vos voisins. Regarder un nouvel objet pourrait vous mettre dans un état de confusion momentané : « C'est ve... – je veux dire *rouge* ! » Et votre crispation face à la lumière verte ? Se manifesterait-elle toujours comme une réponse anormale de votre peau galvanisée ? Pour les besoins de l'argument, le qualophile doit imaginer, aussi peu vraisemblable que cela soit, que *toutes* vos dispositions réactives s'adaptent, ne laissant derrière que le résidu des qualia toujours inversés ; on admettra donc, pour les besoins de l'argument, que la plupart des propensions fondamentales et innées de vos espaces qualitatifs « s'adaptent » aussi – c'est saugrenu, mais le pire reste à venir.

Pour se faire une idée du cadre dont il a besoin, le qualophile doit supposer que toutes ces adaptations deviennent finalement une seconde nature – rapide et spontanée. (Si elles n'étaient pas devenues une seconde nature, il resterait des dispositions réactives qui seraient encore différentes, et l'argumentation exige de les bannir toutes.) Qu'il en soit donc ainsi. À présent, à supposer que *toutes* vos dispositions réactives aient été rétablies, quelle est votre intuition sur les qualia ? Sont-ils toujours inversés ou non ?

Il est légitime de passer à ce point, dans la mesure où après avoir demandé de tolérer tant de présupposés douteux introduits

pour les besoins de l'argument, ou bien on repart les mains vides – aucune intuition ne se fait jour – ou bien on devient très méfiant à l'égard de toute intuition qui pourrait se manifester. Mais peut-être qu'il vous paraît tout à fait évident que vos qualia seraient toujours inversés. Mais pourquoi ? Qu'est-ce qui, dans toute cette histoire, vous a conduit à voir les choses ainsi ? Peut-être que, même si vous avez suivi certaines directions, vous avez en toute innocence ajouté d'autres présupposés que n'appelait pas l'histoire, ou peut-être que vous n'avez pas remarqué certaines possibilités qu'elle ne rejette pas. Je suggère que l'explication la plus légitime de votre intuition que, dans le cas imaginé, vous auriez toujours des « qualia inversés », c'est que vous présupposez aussi, ce qui ne va pas du tout de soi, que l'adaptation se produit entièrement « du côté post-expérientiel ».

Pourtant, il se pourrait que l'adaptation s'effectue sur le chemin ascendant, non ? Lorsque vous mettez pour la première fois vos lunettes fortement teintées, vous ne verrez absolument aucune couleur – ou du moins les couleurs que vous verrez seront étranges et difficiles à distinguer – mais après les avoir un peu portées, vous serez surpris de voir revenir une vision chromatique normale. (Cole, 1990, attire l'attention des philosophes sur ces effets, que vous pouvez tester par vous-mêmes avec les lunettes infrarouges des tireurs embusqués que l'on trouve dans les surplus de l'armée.) Peut-être que, par ignorance de ce fait surprenant, il ne vous est jamais venu à l'esprit que vous *pourriez* vous adapter à l'opération d'une manière très semblable. Nous aurions pu mettre en lumière cette possibilité dans l'expérience de pensée, en ajoutant quelques détails :

... Et au fur et à mesure de l'adaptation, vous vous êtes souvent aperçu, à votre grand étonnement, que les couleurs des choses ne vous paraissaient pas étranges, finalement, et vous vous êtes parfois sentis gênés et avez fait des *doubles* corrections. Lorsque l'on vous a demandé la couleur d'un nouvel objet, vous avez dit « c'est vert – pas rouge – non, c'est vert ! ».

Ainsi racontée, l'histoire pourrait donner l'impression « évidente » que les qualia de couleurs se sont eux-mêmes adaptés, ou ont été réinversés. Mais en tout cas, devez-vous maintenant penser, il faut que ce soit l'un ou l'autre. Il ne pourrait pas y avoir de cas où l'on ne pourrait dire de façon parfaitement évidente quel genre d'ajustement vous avez effectué ! Le présup-

posé naturel qui fonde *cette* conviction est que toutes les adaptations peuvent entrer dans la catégorie soit de la pré-expérience, soit de la post-expérience (stalinienne ou orwellienne). À première vue, cela peut paraître un présupposé innocent, puisque les cas extrêmes sont faciles à classer. Lorsque le cerveau compense pour les mouvements de tête et d'yeux, produisant un champ visuel stable « dans l'expérience », c'est certainement une annulation pré-expérientielle des effets, une adaptation sur le chemin ascendant de la conscience. Et lorsque vous imaginez être en train de faire des compensations périphériques (« tardives ») dans le choix des termes de couleur (« C'est ve... – Je veux dire *rouge* ! »), il s'agit à l'évidence d'un ajustement post-expérientiel, simplement comportemental. N'est-il pas alors raisonnable de penser que lorsque toutes les adaptations ont été faites, ou bien elles laissent la couleur subjective (la couleur « dans la conscience ») inversée ou bien ce n'est pas ce qui se passe ? Voici comment, serions-nous tentés de dire : ajoutons les changements d'aiguillage sur le chemin ascendant : s'il y a un nombre pair – comme dans l'équipe B du travail bien fait – les qualia sont normalisés, et s'il est impair, les qualia restent inversés. C'est absurde. Souvenons-nous de la courbe néo-Laffer du chapitre 4. Il n'y a aucune nécessité logique ou géométrique à ce qu'il y ait une seule et unique valeur de variable discriminée que l'on puisse sélectionner comme étant *la* valeur de la variable « dans la conscience ».

Nous pouvons le démontrer à l'aide d'une petite fantaisie de notre cru, en jouant avec les règles du qualophile. Supposons qu'avant l'opération une certaine nuance de bleu ait eu tendance à vous faire penser à une voiture dans laquelle vous avez eu un jour un accident, et qu'il s'agit donc d'une couleur à éviter. Au premier abord, après l'opération, vous n'avez aucune réaction négative aux choses de cette couleur, vous les trouvez, on peut du moins le supposer, d'un jaune innocent et peu mémorable. Mais après votre complète adaptation, vous rejetez de nouveau les choses qui ont cette nuance de bleu, *et c'est parce qu'elles vous font penser à l'accident*. (Si ce n'était pas le cas, ce serait une disposition réactive inadaptée.) Mais si nous vous demandons si c'est parce que, en vous rappelant l'accident, la voiture était jaune – exactement comme l'objet nuisible qui se trouve maintenant en face de vous – ou si c'est parce que, en vous rappelant l'accident, la voiture était bleue – tout comme l'objet nuisible

qui se trouve maintenant devant vous, vous ne seriez pas vraiment capable de répondre. Votre comportement verbal sera totalement « adapté » ; votre réponse immédiate, de seconde nature, à la question. « De quelle couleur était la voiture avec laquelle vous avez eu un accident ? », c'est « bleue » ; et vous appellerez aussi sans hésiter bleu l'objet nuisible qui se trouve en face de vous. Cela implique-t-il que vous ayez *oublié* la longue période d'entraînement ?

Non. Point n'est besoin de quelque chose d'aussi dramatique que l'amnésie pour expliquer pourquoi vous êtes incapable de répondre, car nous avons des tas de cas quotidiens dans lesquels le même phénomène se produit. Aimez-vous la bière ? Bien des gens qui aiment la bière reconnaîtront que la bière est un goût que l'on acquiert. On s'entraîne progressivement – ou on se met tout simplement – à apprécier cette saveur. Quelle saveur ? La saveur du premier verre ?

Personne ne pourrait aimer *cette* saveur [pourrait répliquer un buveur de bière expérimenté]. La bière a un goût différent pour le buveur de bière expérimenté. Si la bière continuait à avoir pour moi le même goût que celui qu'elle avait à mon premier verre, je n'aurais jamais continué à en boire ! Ou, pour dire la même chose un peu différemment, si mon premier verre de bière avait eu le même goût que celui qu'avait mon tout dernier verre, je n'aurais jamais eu besoin, pour commencer, d'acquérir le goût ! J'aurais adoré le premier verre autant que celui que je viens d'apprécier.

Si ce buveur de bière a raison, alors la bière n'est *pas* un goût que l'on acquiert. Personne ne retrouve dans chaque gorgée *le goût que pouvait avoir le premier verre*. Il faut plutôt dire que *le goût que peut avoir pour eux la bière* change petit à petit. D'autres buveurs de bière pourraient soutenir le contraire : que la bière a bel et bien le même goût aujourd'hui que celui qu'elle a toujours eu ; simplement, aujourd'hui, ils aiment *ce goût même*. Y a-t-il vraiment une différence ? Il y a une différence en hétérophénoménologie, sans aucun doute, et il faut l'expliquer. Il *se pourrait* que les différentes convictions proviennent d'authentiques différences dans la capacité de discrimination du genre : dans la première catégorie de buveur de bière l'« entraînement » a changé la « forme » de l'espace qualitatif gustatif, tandis que dans la seconde l'espace qualitatif reste en gros le même, mais la « fonction d'évaluation » de cet espace a été révisée. Ou bien *il se pourrait* que certains des buveurs de bière ou même tous se

moquent d'eux-mêmes (comme ceux qui soutiennent que les Marilyn à haute résolution sont toutes réellement *là* en arrière-plan de leur champ visuel). Il nous faut regarder au-delà des mondes hétérophénoménologiques et voir ce qui se passe vraiment dans la tête pour voir s'il y a une interprétation (à condition de la « tirer un peu par les cheveux ») qui préserve la vérité des affirmations des buveurs de bière, et *s'il* y en a, ce ne sera que parce que nous décidons de *réduire* « le goût que cela a » à tel ou tel complexe de dispositions réactives (Dennett, 1988). Nous aurions à « détruire » les qualia pour les « rétablir ».

Ainsi, si un buveur de bière fronce les sourcils et affiche une expression d'une gravité mortelle en disant que ce à quoi il fait référence c'est « au goût qu'a aujourd'hui la bière pour moi », il se moque décidément de lui et il pense qu'il peut *par là même* faire référence à un quale qu'il connaît bien, un état subjectif qui est indépendant de ses attitudes réactives changeantes. Il peut avoir l'impression qu'il le peut, mais ce n'est pas vrai ⁵.

Par parité de raisonnement, dans le cas imaginé, où l'objet bleu fait penser à l'accident de voiture, on se moquerait de soi si l'on pensait pouvoir dire, à partir de *la manière dont l'objet vous apparaît*, si c'était « intrinsèquement » la même chose que la manière dont la voiture vous apparaissait au moment de l'accident. Cela suffit pour couper l'herbe sous le pied à l'expérience de pensée du qualophile, car le but était de décrire un cas dans lequel il *était* évident que les qualia seraient inversés alors que les dispositions réactives seraient normalisées. *Présupposer* que l'on pourrait tout simplement le dire est faire une pétition de principe ; or sans ce présumé, il n'y pas d'argument, rien qu'une pompe à intuition – une histoire qui vous caresse dans le sens du poil en vous incitant à mettre en avant vos intuitions viscérales sans leur apporter de bonnes raisons.

Qu'il y ait ou non pétition de principe, il n'empêche qu'il peut *toujours* paraître absolument évident que « les couleurs subjectives dont vous verriez dotées les choses » *devraient* être « comme ceci ou comme cela ». Cela montre tout simplement la puissante force d'attraction que le Théâtre Cartésien exerce sur notre imagination. Cela peut aider à rompre l'attrait résiduel de cette idée si nous considérons un peu plus avant le parallèle proposé avec les lunettes qui inversent les images. Lorsque les adaptations des sujets qui portent ces lunettes sont tellement devenues des secondes natures qu'ils peuvent aller à vélo et à ski, la question naturelle

(mais fourvoyante) à poser est celle-ci : se sont-ils adaptés en *orientant leur univers d'expériences vers le haut*, ou bien en *s'habituant au fait que leur univers d'expériences soit à l'envers* ? Que disent-ils ? Différentes choses, qui sont plus ou moins en corrélation avec le degré d'achèvement de leur adaptation. Plus celle-ci était complète, plus les sujets réputent la question impropre ou prétendent qu'il est impossible d'y répondre. C'est juste ce qu'exige la théorie des Versions Multiples : comme il y a une foule de discriminations et de réactions qui ont besoin d'être ajustées, de se répandre tout autour du cerveau, certaines d'entre elles qui concernent les « réflexes » de bas niveau (tels que baisser la tête dans le bon sens quand quelque chose vous menace) et d'autres qui concernent les actions délibérées sur lesquelles on focalise son attention, il n'est pas surprenant que au fur et à mesure que les adaptations de ce patchwork s'accumulent, les sujets ne savent plus très bien s'il faut dire que « les choses ont l'air qu'elles ont toujours eu » plutôt que « les choses ont toujours l'air différentes, mais je m'y fais ». À certains égards, les choses leur paraissent semblables (à en juger par leurs réactions), et à d'autres égards, elles paraissent différentes (à en juger par d'autres réactions). S'il y avait une seule et unique représentation de l'espace visuello-moteur à travers laquelle toutes les réactions aux stimuli visuels devaient être canalisés, *cela* devrait peut-être apparaître « comme ceci ou comme cela », mais il n'y a pas de représentation unique de ce genre. *La* manière dont les choses leur apparaissent se compose de bon nombre d'habitudes partiellement indépendantes de réaction, et non d'une image unique qui serait située dans la tête et intrinsèquement dirigée de bas en haut, ou de haut en bas. Tout ce qui importe c'est l'accord entre l'entrée et la sortie, et comme cela s'accomplit en maints endroits différents par bien des moyens divers et largement indépendants, on ne peut tout simplement pas dire ce qu'il « faut mettre » sous « mon champ visuel est toujours de haut en bas ».

La même chose vaut pour l'inversion des « qualia ». L'idée qu'il s'agisse de quelque chose *de plus* que l'inversion de toutes les dispositions réactives que l'on peut avoir, de telle sorte que, s'ils étaient renormalisés, les qualia inversés seraient toujours là, relève simplement du mythe tenace du Théâtre Cartésien. Ce mythe est célébré dans les expériences de pensée élaborées sur l'inversion du spectre, mais célébrer n'est pas étayer ni démontrer. Si les qualia ne sont rien de plus que la somme totale des

dispositions à réagir, l'idée consistant à garder constants les qualia tout en ajustant les dispositions est contradictoire.

Des qualia « épiphénoménaux » ?

Une autre expérience philosophique de pensée à propos de notre expérience de la couleur s'est révélée irrésistible : le cas très discuté, évoqué par Frank Jackson (1982), de Marie, spécialiste des couleurs et qui n'en a jamais vu. Comme toute bonne expérience de pensée, ce qu'elle cherche à montrer est tout de suite évident même au non-initié. En fait, c'est une mauvaise expérience de pensée, une pompe à intuition qui nous encourage vraiment à mal comprendre ses prémisses !

Marie est une femme de science brillante qui, pour une raison quelconque, est forcée de faire des recherches sur le monde depuis une chambre noire et blanche et à travers un écran de télévision en noir et blanc. Elle se spécialise dans la neurophysiologie de la vision et acquiert, supposera-t-on, toutes les informations physiques que l'on peut avoir sur ce qui se passe quand nous voyons des tomates mûres ou sur le ciel ; elle utilise des termes tels que « rouge », « bleu », etc. Elle découvre, par exemple, quelles sont exactement les combinaisons de longueurs d'onde en provenance du ciel qui stimulent la rétine, et de quelle manière exactement cela produit à travers le système nerveux central la contraction des cordes vocales et l'expulsion de l'air des poumons qui ont pour résultat l'énoncé de la phrase « Le ciel est bleu »... Que se passera-t-il quand on libérera Marie de son écran de télévision en noir et blanc et qu'on lui donnera un écran de télévision en couleur ? Apprendra-t-elle quelque chose ou pas ? Il semble tout bonnement évident qu'elle apprendra quelque chose sur le monde et sur l'expérience visuelle que nous en avons. Mais il faudra dire alors que sa connaissance antérieure était incomplète. Or elle avait *toutes* les informations physiques. *Ergo*, il faut plus que cela, et le physicalisme est faux...

On ne pourrait guère montrer les choses plus clairement. Marie n'a eu absolument *aucune* expérience de la couleur (il n'y a pas de glace dans laquelle elle puisse se regarder, elle est obligée de porter des gants noirs, etc.) et donc, au moment précis où ses capteurs la laissent finalement entrer dans le monde coloré qu'elle ne connaît que par description (et par des diagrammes en noir et blanc), « il semble tout bonnement évident », comme le dit

Jackson, qu'elle apprendra quelque chose. En vérité, nous pouvons tous aisément l'imaginer en train de voir sa première rose rouge et de s'exclamer : « C'est donc à *cela* que ressemble une rose ! » Et il peut aussi nous venir à l'esprit que si les premières choses colorées qu'on lui montre sont, disons, des blocs de bois sans étiquettes, et qu'on lui dit seulement que l'un d'eux est rouge et l'autre bleu, elle sera totalement incapable de dire duquel il s'agit, jusqu'à ce que d'une manière ou d'une autre elle apprenne quels sont les termes de couleur qui cadrent avec ses toutes nouvelles expériences.

C'est ainsi que presque tout le monde imagine sa première expérience de pensée – pas simplement les non-initiés, mais les philosophes les plus astucieux, les plus blindés (Tye, 1986 ; Lewis, 1988 ; Loar, 1990 ; Lycan, 1990 ; Nemirov, 1990 ; Harman, 1990 ; Block, 1990 ; van Gulick, 1990). Seul Paul Churchland (1985, 1990) a sérieusement résisté à l'*image* si nettement évoquée par l'expérience de pensée. L'image est fausse ; si c'est ainsi que vous vous imaginez le cas, vous ne suivez tout simplement pas les directions ! La raison pour laquelle personne ne suit les directions, c'est que ce qu'ils vous demandent d'imaginer présuppose tant de choses que l'on ne peut même pas essayer. La prémisse cruciale est qu'« elle a *toutes* les informations physiques ». Ce n'est manifestement pas imaginable, aussi personne ne s'en préoccupe-t-il. On imagine simplement qu'elle sait des tas et des tas de choses – peut-être imagine-t-on qu'elle connaît tout ce qu'on sait *aujourd'hui* sur la neurophysiologie de la vision chromatique. Mais c'est une goutte d'eau dans l'océan, et sans doute Marie apprendrait-elle quelque chose si c'était *là* tout ce qu'elle connaissait.

Pour faire ressortir ici l'illusion à laquelle l'imagination est en proie, continuons l'histoire d'une manière surprenante – mais légitime :

Un beau jour, les geôliers de Marie ont décidé qu'il était temps qu'elle voie les couleurs. Pour lui faire un tour, ils ont préparé une banane bleu clair pour la présenter comme sa toute première expérience de couleur. Marie a jeté un coup d'œil et a dit : « Hé là ! Vous avez essayé de me jouer un tour ! Les bananes sont jaunes, or celle-ci est bleue ! » Ses geôliers sont restés médusés. Comment a-t-elle fait ? « C'est bien simple, a-t-elle répondu. Vous devez vous rappeler que je connais *tout* – absolument tout – ce qui peut être connu sur les causes et effets physiques de la vision chro-

matique. Et donc bien sûr avant que vous n'apportiez la banane, j'avais déjà écrit, en détaillant par le menu, quelle impression physique ferait exactement un objet jaune ou un objet bleu (ou un objet vert, etc.) sur mon système nerveux. Je savais donc déjà exactement quelles *pensées* j'aurais (parce que, après tout, la "simple disposition" à penser à ceci ou à cela n'est pas l'un de vos fameux *qualia*, si ?). Je n'ai pas été le moins du monde surprise par mon expérience de bleu (ce qui m'a surprise c'est que vous m'ayez joué un tour aussi médiocre). Je me rends compte que *c'est difficile pour vous d'imaginer* que je puisse en savoir tant sur mes dispositions réactives que la manière dont le bleu m'a affectée n'ait pas été pour moi une surprise. Bien sûr, c'est difficile pour vous à imaginer. C'est difficile pour n'importe qui d'imaginer les conséquences que tirera quelqu'un qui connaît absolument tout sur le physique de quelque chose que ce soit ! »

J'ai évidemment triché, pensez-vous. Je cache sans doute telle ou telle impossibilité derrière le voile des remarques de Marie. Pouvez-vous le prouver ? Ce que je cherche à montrer ce n'est pas que ma façon de raconter le reste de l'histoire prouve que Marie *n'apprend* pas quelque chose, mais que la manière habituelle d'imaginer l'histoire ne *prouve* pas qu'elle le *fasse*. Cela ne prouve rien ; cela pompe simplement l'intuition que c'est ce qu'elle fait (« cela semble tout bonnement évident ») en nous berçant et en nous conduisant à imaginer quelque chose d'autre que ce qui est requis par les prémisses.

Il est bien sûr vrai que dans n'importe quelle version réaliste, aisément imaginable, elle pourrait en savoir beaucoup, mais elle ne saurait pas tout ce qui touche au physique. Se contenter d'imaginer que Marie en sait beaucoup et en rester là n'est pas une bonne façon de se représenter ce qu'implique le fait pour elle de posséder « toutes les informations physiques » — pas plus que d'imaginer qu'elle est monstrueusement riche ne serait une bonne façon de se représenter ce qu'implique pour elle l'hypothèse qu'elle possède tout. Cela peut nous aider d'imaginer l'étendue des pouvoirs que son savoir lui donne si nous commençons à énumérer quelques-unes des choses qu'elle sait à l'évidence par avance. Elle connaît le noir et le blanc, et les nuances de gris, et elle peut faire la différence entre la couleur de n'importe quel objet et de telles propriétés de surface telles que le brillant par opposition au mat, et elle sait tout de la différence qui existe entre des frontières de luminosité et des frontières de couleur

(les frontières de luminosité sont celles qui se manifestent sur une télévision en noir et blanc, pour dire les choses grossièrement). Et elle sait précisément quels effets – décrits en termes neurophysiologiques – telle ou telle couleur donnée aura sur son système nerveux. La seule tâche qui lui reste donc c'est de concevoir un moyen d'identifier « de l'intérieur » ces effets neurophysiologiques. Il se peut que vous puissiez nettement imaginer qu'elle fait sur ce point *un léger* progrès – par exemple en concevant des manières astucieuses qui lui permettent de dire que telle couleur, quelle qu'elle soit, n'est *pas* jaune, ou n'est *pas* rouge. Comment ? En notant telle ou telle réaction saillante et spécifique qu'aurait son cerveau pour le jaune, ou seulement pour le rouge. Mais si vous lui laissez mettre ainsi ne serait-ce qu'un pied dans son espace chromatique, il vous faudra conclure alors qu'elle peut chercher à faire du chemin pour compléter ses progrès en matière de connaissance, parce qu'elle ne connaît pas simplement les réactions *saillantes*, elle les connaît toutes.

Souvenons-nous de la boîte en carton de Julius et Ethel Rosenberg, qu'ils ont transformée en un détecteur M. Imaginons à présent leur surprise si un imposteur arrivait avec un morceau « qui s'adapte » mais qui ne soit pas l'original. « Impossible ! » s'écrient-ils. « Pas impossible, dit l'imposteur, simplement difficile. J'avais *toutes les informations* requises pour reconstruire un détecteur M et pour faire une autre chose avec la propriété formelle M. » Marie avait assez d'information (dans le cas original, correctement imaginé) pour concevoir tout simplement ce qu'étaient ses détecteurs de rouge et ses détecteurs de bleu, et donc pour les identifier à l'avance. Elle ne procédait pas comme on apprend habituellement les couleurs, mais Marie n'a rien d'une personne habituelle.

Je sais que cela ne satisfera pas un certain nombre des fans philosophiques de Marie, et qu'il y a beaucoup plus à dire, mais – et c'est ce que je voudrais surtout dire – la véritable démonstration doit se poursuivre dans un domaine très éloigné de l'exemple de Jackson, qui est une manière classique de provoquer le syndrome des Philosophes : prendre à tort un échec de l'imagination pour l'intuition d'une nécessité. Certains des philosophes qui ont traité du cas de Marie peuvent ne pas se soucier du fait que ce qu'ils ont imaginé est faux, puisqu'ils s'en sont simplement servis comme d'un tremplin pour des discussions qui ont éclairé diverses questions en elles-mêmes intéressantes et importantes.

Je n'irai pas plus loin ici sur ces sujets, puisque ce qui m'intéresse, c'est de considérer directement la conclusion que tire lui-même Jackson de son exemple : les expériences visuelles ont des qualia qui sont « épiphénoménaux ».

Le terme d'« épiphénomène » est d'usage courant aujourd'hui chez les philosophes comme chez les psychologues (et autres cognitivistes). On l'utilise en présupposant que sa signification est familière et qu'elle fait l'objet d'un accord, alors qu'en fait, les philosophes et les cognitivistes emploient ce terme en des sens *totale*ment différents – fait étrange, et qui me paraît encore plus étrange bien que je l'aie fait observer à maintes reprises, personne ne semble s'en soucier. Comme l'« épiphénoménisme » apparaît souvent comme le dernier refuge des qualia, et que cette apparence de sécurité est entièrement due à la confusion entre ces deux significations, je dois me faire l'avocat du diable et mettre en garde ceux qui emploient ce terme.

Selon le *Shorter Oxford English Dictionary*, le terme « épiphénomène » apparaît pour la première fois en 1706 comme terme de pathologie, « un symptôme en apparence secondaire ». Le biologiste évolutionniste Thomas Huxley (1874) fut probablement l'auteur qui a étendu le terme à son usage courant en psychologie, où il signifie une propriété ou un sous-produit non fonctionnels. Huxley employait ce terme dans sa discussion de l'évolution de la conscience où il soutenait que les propriétés épiphénoménales (comme « le sifflement de la machine à vapeur ») ne pouvaient s'expliquer par la sélection naturelle.

Voici un exemple clair de son usage du mot :

Pourquoi les gens qui sont plongés dans une profonde réflexion se mordent-ils les lèvres et tapent-ils du pied ? Est-ce que ces actions ne sont que des épiphénomènes qui accompagnent le déroulement central des sentiments et de la pensée ou bien se pourrait-il qu'elles fassent elles-mêmes partie intégrante de ce déroulement ?

On notera que les auteurs entendent affirmer que ces actions, bien qu'on puisse parfaitement les entrevoir, ne jouent aucun rôle opératoire, aucun rôle structurel dans le déroulement des sentiments et de la pensée ; elles sont épiphénoménales. Dans le même esprit, le ronronnement de l'ordinateur est épiphénoménal, comme l'est votre ombre quand vous vous faites une tasse de thé. Les épiphénomènes sont de simples sous-produits, mais comme tels ils sont des produits qui ont de nombreux effets dans

le monde ; taper du pied fait un bruit qu'on peut enregistrer ; et votre ombre a ses effets sur un film photographique, pour ne rien dire du léger refroidissement des surfaces sur lesquelles elle se répand.

Le sens philosophique standard est différent : « x est épiphénoménal » veut dire que « x est un effet mais n'a pas lui-même le moindre effet dans le monde » (voir Broad, 1925, p. 118, pour la définition qui inaugure, ou du moins établit l'usage philosophique). Ces significations sont-elles réellement si différentes ? Oui, aussi différentes que les significations de meurtre et de mort. La signification philosophique est plus forte : tout ce qui n'a pas le moindre effet dans le monde physique n'a sûrement aucun effet sur la fonction de quoi que ce soit, mais la converse ne s'ensuit pas, comme le montre bien l'exemple de Zajonc et de Markus.

En fait, la signification philosophique est trop forte ; elle donne un concept qui est totalement inutile (Harman, 1988 ; Fox, 1989). Comme x n'a pas d'effets physiques (selon cette définition), aucun instrument ne peut entrevoir directement ou indirectement la présence de x ; le cours des choses dans le monde n'est en aucune façon modulé par la présence ou l'absence de x . Comment pourrait-il donc y avoir la moindre raison empirique d'affirmer la présence de x ? Supposons, par exemple, que Otto insiste sur le fait que lui (au moins) a des qualia épiphénoménaux. Pourquoi dit-il cela ? Non parce qu'ils ont un effet sur lui, qui le guident ou l'alertent en un sens quand il fait ses déclarations. Par la définition même des épiphénomènes (au sens philosophique), les déclarations qui viennent du fond du cœur d'Otto, selon lesquelles il a des épiphénomènes, *ne pourraient pas* pour lui ou pour qui que ce soit d'autre être une preuve qu'il les a, puisqu'il dirait exactement la même chose s'il ne les avait pas. Mais peut-être qu'Otto a des preuves « internes » ?

Il y a ici une échappatoire, mais qui n'est pas très attirante. On a défini les épiphénomènes, souvenez-vous, comme n'ayant aucun effet dans le monde *physique*. Si Otto veut épouser un dualisme intégral, il peut soutenir que ses qualia épiphénoménaux n'ont pas d'effets dans le monde physique, mais ont des effets dans son monde mental (non physique ; Broad, 1925, a fermé la porte, par définition, à cette échappatoire, mais on est libre d'en faire la demande). Par exemple, *ils causent certaines de ses croyances (non physiques)*, telles que sa croyance qu'il a

des qualia épiphénoménaux. Mais ce n'est qu'une manière temporaire de se sortir d'embarras. Car voici qu'à leur tour, sous peine de contradiction, ses croyances ne peuvent avoir aucun effet dans le monde physique. S'il perdait brusquement des qualia épiphénoménaux, il ne croirait plus qu'il les a, mais il continuerait toujours à *dire* que si. Il ne croirait tout simplement pas ce qu'il est en train de dire ! (Pas plus qu'il ne pourrait vous dire qu'il ne croyait pas ce qu'il disait, ou faisait quelque chose qui révélait qu'il ne croyait plus ce qu'il disait.) La seule manière donc dont Otto puisse « justifier » sa croyance dans les épiphénomènes serait de battre en retraite dans un monde solipsiste où il n'y aurait que lui, ses croyances et ses qualia, coupés de tous les effets qu'il peut y avoir dans le monde. Loin d'être une manière « sûre » d'être matérialiste tout en gardant vos qualia, c'est au mieux une manière d'épouser le solipsisme le plus radical, en coupant votre esprit – vos croyances et vos expériences – de tout commerce avec le monde matériel.

Si les qualia sont épiphénoménaux, au sens philosophique standard, leur occurrence ne peut expliquer la manière dont les choses se produisent (dans le monde matériel) puisque, par définition, les choses se produiraient exactement de la même manière sans eux. Il ne pourrait donc y avoir de raison empirique de croire aux épiphénomènes. Pourrait-il y avoir une autre sorte de raison d'affirmer leur existence ? Quelle sorte de raison ? Une raison *a priori*, probablement. Mais laquelle ? Personne n'en a jamais proposé – bonne, mauvaise, indifférente – que j'aie pu voir. Si quelqu'un veut objecter que je suis un « vérificationniste » au sujet de ces épiphénomènes, je réponds : tout le monde n'est-il pas vérificationniste sur ce genre d'assertion ? Considérons, par exemple, l'hypothèse qu'il y a quatorze gremlins épiphénoménaux dans chacun des cylindres d'un moteur à combustion interne. Ces gremlins n'ont aucune masse, aucune énergie, aucune propriété physique ; ils ne font pas tourner le moteur plus doucement ou plus fort, plus vite ou plus lentement. Il n'y a *et ne pourrait y avoir* aucune preuve de leur présence, aucun moyen empirique en principe de distinguer cette hypothèse de ses rivales : il y a douze ou treize ou quinze... gremlins. Par quel principe défend-on le rejet en bloc de telles absurdités ? Un principe vérificationniste ou le pur et simple bon sens ?

Ah, mais il y a une différence ! dit Otto. Il n'y a aucune raison indépendante de prendre l'hypothèse de ces gremlins au sérieux. Tu les as juste fabriqués sous l'inspiration du moment. Les qualia, en revanche, sont là depuis bien longtemps, et jouent un rôle majeur dans notre schème conceptuel !

Que se passerait-il alors si certains attardés pensaient depuis des générations que les gremlins ont fait rouler leur voiture, et se voyaient à présent réduits en raison du progrès scientifique à soutenir la thèse désespérée que les gremlins sont bien là, soit, mais sont épiphénoménaux ? Est-ce pour nous une erreur de rejeter leur « hypothèse » en ruant dans les brancards ? Quel que soit le principe sur lequel nous nous appuyons, lorsque nous éconduisons du revers de la main de telles absurdités, il suffit de rejeter la doctrine qui veut que les qualia soient épiphénoménaux en ce sens philosophique. Ce ne sont pas des thèses qui méritent d'être discutées en s'y attaquant de face.

On a du mal à croire que les philosophes qui ont récemment décrit leurs thèses comme de l'épiphénoménisme puissent faire une erreur aussi navrante. Peut-être se contentent-ils d'affirmer que les qualia sont épiphénoménaux au sens de Huxley ? Les qualia, selon cette interprétation, *sont* des effets physiques et *ont* des effets physiques ; ils ne sont pas simplement fonctionnels. N'importe quel matérialiste serait heureux d'admettre que son hypothèse est vraie – si nous identifions les qualia à des dispositions réactives, par exemple. Comme on l'a noté dans la discussion sur ce que l'on apprécie, même si certaines déformations ou tendances dans nos espaces qualitatifs sont fonctionnelles – ou étaient fonctionnelles – d'autres sont simplement le fait du hasard. Pourquoi est-ce que je n'aime pas les brocolis ? Probablement pour aucune espèce de raison ; ma disposition réactive négative est purement épiphénoménale, un sous-produit de mon câblage, sans signification aucune. Elle n'a aucune fonction, mais a énormément d'effets. Dans n'importe quel système structurellement organisé, certaines propriétés sont cruciales alors que d'autres sont plus ou moins révisables *ad libitum*. Tout doit se faire de telle ou telle manière, mais souvent les manières elles-mêmes importent peu. Le levier de vitesses sur ma voiture peut avoir une certaine longueur et une certaine force, mais qu'il soit de coupe ronde, carrée ou ovale est une propriété épiphénoménale, au sens de Huxley. Dans les systèmes CADAVEUGLES que nous avons imaginés au chapitre 9, le schème chromatico-numé-

rique de codage particulier que nous avons imaginé était épiphénoménal. Nous pouvions « l'inverser » (en utilisant des nombres négatifs, ou en multipliant toutes les valeurs par une certaine constante) sans introduire la moindre différence *fonctionnelle* quant aux prouesses qu'elle accomplit dans le traitement d'information. Une telle inversion pourrait fort bien être indécélable à l'inspection ordinaire, et indécélable *par le système*, mais elle ne serait pas épiphénoménale au sens philosophique. Il y aurait, par exemple, de nombreuses infimes différences de voltage dans les registres mémoriels contenant les différents nombres.

Si nous pensons à toutes les propriétés de notre système nerveux qui nous permettent de voir, d'entendre, de sentir, de goûter et de toucher les choses, nous pouvons, en gros, les diviser en propriétés qui jouent vraiment les rôles cruciaux en transportant les traitements d'information et les propriétés épiphénoménales qui sont plus ou moins révisables *ad lib.* comme le système de codage chromatique du système CADAVEUGLES. Lorsqu'un philosophe suppose que les qualia sont des propriétés épiphénoménales d'états cérébraux, cela pourrait vouloir dire que les qualia pourraient se révéler être des variations locales de la chaleur générées par le métabolisme neuronal. Ce n'est sûrement pas ce que les épiphénoménistes ont à l'esprit, si ? Si oui, alors les qualia comme épiphénomènes ne sont pas une objection au matérialisme. Le temps est venu de laisser carrément la charge de la preuve à ceux qui persistent à employer le terme. Le sens philosophique du terme est tout bonnement ridicule ; le sens de Huxley est relativement clair et non problématique – et peu pertinent pour les arguments philosophiques. Aucun autre sens du terme n'a la moindre valeur. Si donc quelqu'un prétend soutenir une variété d'épiphénoménisme, tâchez de rester poli, mais demandez-lui de quoi il parle *au juste*.

Notez, au passage, que cette équivoque entre deux sens du terme « épiphénoménal » empoisonne aussi la discussion sur les zombies. Il est impossible, souvenez-vous, de distinguer par le comportement un zombie de philosophe d'un être humain normal, mais le premier n'est pas conscient. Il n'y a rien qui ressemble à l'effet que cela fait d'être un zombie ; cela paraît juste être ainsi aux yeux des observateurs (y compris à ses propres yeux, comme on l'a vu au chapitre précédent). Mais on peut donner à cela une interprétation faible ou forte, selon la manière

de traiter ce fait qu'il soit impossible pour les observateurs de faire la distinction. Si nous devions déclarer que *en principe* il est impossible de distinguer un zombie d'une personne consciente, alors nous dirions que la véritable conscience est épiphénoménale *au sens ridicule*. C'est tout simplement idiot. Nous pourrions donc dire plutôt que la conscience pourrait être épiphénoménale, au sens de Huxley : bien qu'il y eût une manière de distinguer les zombies de personnes réelles (qui sait, peut-être que les zombies ont un cerveau vert), la différence ne se manifeste pas comme une différence fonctionnelle *aux observateurs*. De même, les corps humains qui ont un cerveau vert n'abritent pas d'observateurs, alors que c'est le cas d'autres corps humains. D'après cette hypothèse, nous serions en principe capables de distinguer les corps habités des corps inhabités en contrôlant la couleur de leur cerveau. C'est également idiot, bien sûr, et dangereusement idiot, car cela fait écho au genre de préjugés entièrement immotivés qui ont fait refuser à des gens le plein statut de personne sur la base de la couleur de leur peau. Il est temps de reconnaître l'idée de la possibilité de zombies pour ce qu'elle est : non pas une idée philosophique sérieuse, mais une relique saugrenue et ignoble d'anciens préjugés. Peut-être que les femmes ne sont pas vraiment conscientes ! Les juifs non plus, peut-être ! Autant d'absurdités pernicieuses. Comme le dit Shylock, qui attire notre attention, à fort juste titre, sur les critères « purement comportementaux » :

« Un juif n'a-t-il pas des yeux ? Un juif n'a-t-il pas des mains, des organes, des proportions, des sens, des affections, des passions ? N'est-il pas nourri de la même nourriture, blessé des mêmes armes, sujet aux mêmes maladies, guéri par les mêmes moyens, échauffé et refroidi par le même été et par le même hiver qu'un chrétien ? Si vous nous piquez, est-ce que nous ne saignons pas ? Si vous nous chatouillez, est-ce que nous ne rions pas ? Si vous nous empoisonnez, est-ce que nous ne mourons pas ? » (trad. François-Victor Hugo, Gallimard, La Pléiade, vol. 1, p. 1234-1235).

Il y a une autre manière de concevoir la possibilité de zombies, et à certains égards, je pense qu'elle est plus satisfaisante. Est-ce que les zombies sont possibles ? Ils ne sont pas simplement possibles, ils sont réels. Nous sommes tous des zombies⁶. Personne n'est conscient — pas de la manière systématiquement mystérieuse qui renforce des doctrines telles que l'épiphénoménalisme ! Je ne puis prouver qu'aucune conscience de ce genre n'existe pas. Je ne puis également prouver que les gremlins

n'existent pas. Le mieux que je puisse faire c'est de montrer qu'il n'y a aucune raison sérieuse d'y croire.

Pour en revenir à mon rocking-chair

Au chapitre 1, deuxième section, j'ai entrepris d'expliquer la conscience en rappelant un épisode de ma propre expérience consciente alors qu'assis sur mon rocking-chair, je regardais par la fenêtre, un jour radieux de printemps. Retournons à ce passage et voyons comme la théorie que j'ai développée en traite. Voici le texte :

Un rayon de soleil mordoré pénétrait par la fenêtre en ce début de printemps, et dans le jardin les milliers de branches et de ramures de l'érable étaient encore bien visibles à travers une brume de bourgeons verts formant une élégante trame merveilleuse de complexité. La vitre est en verre ancien, et il y a dedans une ride qu'on décèle à peine. Au fur et à mesure de mon balancement sur la chaise, cette imperfection du verre produisit une vague de tressautements synchronisés qui allaient et venaient à travers le delta de branches. Ce mouvement régulier se superposait avec une vivacité remarquable au chatoiement plus chaotique des ramures et des branchages qu'ébranlait la brise. C'est alors que je remarquai que ce métronome visuel des branches de l'arbre suivait étroitement le rythme du concerto grosso de Vivaldi que j'étais en train d'écouter en « musique d'arrière-fond » pendant ma lecture... Ma pensée consciente, en particulier le plaisir que je prenais au mélange de rayons de soleil, de violons vivaldiens ensoleillés, d'ondulations des branches, le plaisir que je prenais à penser à tout cela, comment *tout cela* pouvait-il être seulement quelque chose de physique qui se produisait dans mon cerveau ? Comment une combinaison d'événements électrochimiques ayant lieu dans mon cerveau pouvait-elle en quelque sorte s'ajouter à la façon délicieuse dont des centaines de bourgeons se balançaient en mesure avec la musique ? Comment un processus de traitement de l'information localisé dans mon cerveau pouvait-il *être* la délicate douceur de la lumière du soleil que je sentais tomber sur moi ?... Cela semble vraiment impossible.

Comme je nous ai tous encouragés à nous conduire en hétérophénoménologues, je vois mal comment je pourrais me dispenser d'être aussi bien sujet que praticien et ne pas en être heureux. Voyons donc comment cela se passe : j'applique ma

propre théorie à moi-même. En tant qu'hétérophénoménologues, notre tâche est de prendre ce texte, de l'interpréter, et de mettre alors en relation les objets du monde hétérophénoménologique dennettien qui en sortent avec les événements qui se produisent à ce moment-là dans le cerveau de Dennett.

Comme le texte a été produit il y a des semaines ou des mois après que les événements dont il s'agit se sont produits, nous pouvons être sûrs qu'il a été abrégé, non seulement par les compressions éditoriales délibérées de l'auteur, mais aussi par les processus inexorables d'abréviation qu'effectue la mémoire à travers le temps. Si nous avions essayé de le faire plus tôt – si l'auteur s'était saisi d'un magnétophone au moment où il était assis sur son rocking-chair, et avait produit de temps à autre le texte – il aurait été à coup sûr très différent. Non seulement plus riche dans le détail, et plus désordonné, mais aussi, bien sûr, refaçonné et réorienté par les propres réactions de l'auteur à la création du texte – lequel auteur eût été à l'écoute des sons réels de ses propres mots au lieu d'être en train d'y songer vaguement en silence. Parler à haute voix, comme le sait bien tout conférencier, révèle souvent des implications (et particulièrement des problèmes) dans le message que l'on est en train de délivrer, qui vous échappent lorsque vous vous livrez à un soliloque silencieux.

Tel qu'il est, le texte décrit une portion simple (et sans nul doute idéalisée) des contenus de la conscience de l'auteur. Mais nous devons faire attention de ne pas supposer que les « parties laissées de côté » dans le texte donné étaient toutes « réellement présentes » dans quelque chose que nous pourrions appeler le cours de la conscience de l'auteur. Nous ne devons pas commettre l'erreur qui consiste à supposer qu'il y a des faits – des faits qu'on ne peut pas recouvrer mais qui sont réels – permettant de dire quels contenus au juste étaient conscients et quels contenus ne l'étaient pas alors. Et en particulier, nous ne devons pas supposer que lorsqu'il a regardé par la fenêtre, il « a tout absorbé » en n'en faisant qu'une seule et merveilleuse bouchée mentale – même si c'est ce que son texte décrit. Il lui a semblé, d'après le texte, que son esprit – son champ visuel – était rempli de détails complexes de bourgeons vert et or et de branches ondulantes, mais même si c'est ainsi que les choses lui paraissent être, c'était une illusion. Aucun « plein » de ce genre ne lui est jamais venu à l'esprit ; le plein est resté dehors dans le monde-là où il n'avait pas à être *représenté*, mais pouvait simplement *être*. Quand nous

nous émerveillons, dans ces moments de conscience de soi avivée, de la glorieuse richesse de notre expérience consciente, la richesse dont nous nous émerveillons est vraiment la richesse du monde au-dehors, dans tous ses détails séduisants. Elle n'entre pas dans notre esprit conscient, mais est simplement là à portée de main.

Que dire de toutes les branches et brindilles ondulant à l'unisson ? Les branches qui sont là dehors sur l'arbre n'ont pas ondulé, c'est certain, parce que l'ondulation était due aux plissements de la vitre ; ce qui ne veut pas dire que toute cette ondulation devait se produire dans l'esprit ou le cerveau de l'auteur tout comme elle s'est produite à l'intérieur du carreau qui en a été la cause. Si quelqu'un avait filmé le changement d'images sur la rétine de l'auteur, il y aurait trouvé l'ondulation, tout comme dans un film, mais il n'y a guère de doute sur l'endroit où toute l'ondulation ou presque s'est arrêtée ; ce qui s'est passé à l'intérieur de sa rétine est simplement qu'il a reconnu où il y avait, comme il le dit dans le texte, une merveilleuse vague d'ondulations synchronisées dont il devait faire l'expérience. Il a vu les ondulations, et il a vu jusqu'où elles s'étendaient, tout comme vous verriez toutes les Marilyns sur le papier peint. Et comme sa rétine était pourvue d'une bonne dose d'ondulation, s'il avait eu envie de poursuivre l'expérience, il y *aurait eu* encore plus de détails dans les multiples versions dont notre texte est tout ce qui reste.

Il y avait bien d'autres détails sur lesquels l'auteur aurait pu mettre l'accent, mais il ne l'a pas fait. Il y a plein de faits décisifs authentiques mais irrécupérables qui permettraient de dire quels sont les détails qui ont été ici ou là discriminés par divers systèmes de son cerveau, mais la somme totale de ces faits décisifs ne permet pas d'établir desquels de ces faits décisifs il était décidément et réellement conscient (mais qu'il a oublié au moment où il a produit son texte), et qui étaient décidément et vraiment à l'« arrière-plan » de sa conscience (bien qu'il n'y ait pas prêté alors attention). Notre tendance à supposer qu'il *doit* y avoir un fait décisif qui établisse de telles questions revient à supposer, tel le lecteur naïf, qu'il *doit* y avoir une réponse à des questions telles que : Est-ce que Sherlock Holmes a eu des œufs à son petit déjeuner le jour où le Dr Watson l'a rencontré ? Conan Doyle *aurait pu* mettre ce détail dans le texte, mais il ne l'a pas fait, et puisqu'il ne l'a pas fait, il n'y a tout simplement aucun fait décisif qui permette de dire si ces œufs relèvent *du monde*

fictionnel de Sherlock Holmes. Même si Conan Doyle *pensait* à Holmes en train de manger des œufs ce matin-là, même si, dans une version antérieure, Holmes *est représenté en mots manuscrits* comme en train de manger des œufs ce matin-là, il n'y a tout simplement aucun fait décisif qui permette de dire si dans l'univers fictionnel de Sherlock Holmes, l'univers constitué à partir du texte publié que nous avons vraiment, il a eu des œufs au petit déjeuner.

Le texte que nous avons de Dennett n'a pas été « écrit dans son cerveau » entre le moment où ce dernier était dans son rocking-chair et le moment où le texte a été tapé et classé dans un dossier de son traitement de textes. L'attention qu'il a manifestée lorsqu'il était dans son rocking-chair, et la reconstitution concomitante de ces particuliers qui ont attiré son attention, ont eu pour effet de fixer les contenus de ces particuliers de façon relativement sûre « en mémoire », mais il ne faudrait pas considérer cet effet comme le stockage d'une image (ou d'une phrase) ou de toute autre représentation saillante. On devrait plutôt y penser comme à quelque chose qui rend une récurrence partiellement similaire de l'activité plus vraisemblable, et cet événement vraisemblable est ce qui s'est produit, peut-on supposer, au moment où l'on a tapé à la machine, incitant les démons des mots qui s'agitent dans son cerveau à se former en des coalitions qui ont produit, pour la première fois, un enchaînement de phrases. Or ce qui s'est en partie produit plus tôt, sur le rocking-chair, contenait sans nul doute des mots et expressions de véritable anglais, et cette collaboration première entre des contenus sans mots et des mots a sans nul doute facilité la récupération de certaines des expressions anglaises mêmes, quand est venu le moment de taper à la machine.

Revenons à l'univers hétérophénoménologique de ce texte. Que dire de la joie qu'il évoque ? « ... La combinaison de la lumière du soleil, les violons ensoleillés de Vivaldi, les branches qui ondulent – plus le plaisir que j'ai eu à penser à tout cela... » On ne pourrait expliquer cela par l'invocation de qualia intrinsèquement plaisants de la vue, du son, et de la pure et simple pensée. L'idée qu'il existe de tels qualia nous écarte simplement de toutes les voies possibles d'explication, captivant notre attention, à la manière dont un doigt qui se balance devant les yeux d'un bébé peut captiver son attention, nous faisant écarquiller bêtement les yeux devant l'« objet intrinsèque » alors que nous

devrions nous mettre en quête d'une description des mécanismes sous-jacents *et* d'une explication (finalement une explication évolutionniste) des raisons pour lesquelles les mécanismes font ce qu'ils font.

Le plaisir de l'auteur s'explique aisément par le fait que toute expérience visuelle est composée par les activités des circuits neuronaux dont l'activité même nous est, de façon innée, agréable, pas seulement parce que nous aimons simplement être informés, mais parce que nous aimons les manières particulières dont nous parvenons à être informés. Le fait que la vue des taches que fait la lumière du soleil sur des bourgeons printaniers soit quelque chose que l'être humain aime n'est pas surprenant. Le fait que certains êtres humains aiment aussi regarder des lamelles microscopiques de bactéries et que d'autres aiment regarder des photographies d'accidents d'avion est plus étrange, mais les sublimations et perversions du désir proviennent des mêmes sources animales dans le câblage de notre système nerveux.

L'auteur continue à se demander comment sur terre « tout ceci pourrait n'être en vérité qu'une combinaison d'événements électrochimiques de mon cerveau ». Comme son étonnement le montre clairement, il ne paraît pas en être ainsi. Ou bien dans tout événement il y a eu un moment où il lui est venu à l'esprit que cela ne lui paraissait pas être simplement une combinaison d'événements électrochimiques dans son cerveau. Mais nos chapitres suivants envisagent une réplique : Eh bien, à quoi pensez-vous que cela ressemblerait si *c'était* juste une combinaison d'événements électrochimiques dans votre cerveau ? Ne nous sommes-nous pas donné de bonnes raisons de conclure qu'avec un cerveau organisé comme l'est le nôtre, c'est juste le genre d'univers hétérophénoménologique auquel nous nous attendrions ? Pourquoi de telles combinaisons d'événements électrochimiques dans le cerveau n'auraient-elles pas précisément les effets que nous avons entrepris d'expliquer ?

(C'est l'auteur qui parle) : Il reste encore une énigme, malgré tout. Comment est-ce que *j'en* viens à savoir tout cela ? Comment est-ce que *je* peux en arriver à vous dire tout ce qui se passait dans ma tête ? La réponse de l'énigme est simple : *parce que c'est ce que je suis*. Parce que quelqu'un qui sait et rapporte ces choses en de tels termes est ce qui est moi. *Mon* existence s'explique par le fait qu'il y a ces capacités dans ce corps.

Cette idée, l'idée d'un moi comme étant le Centre de Gravité

Narrative, est celle que nous sommes finalement prêts à examiner. C'est certainement une idée pour laquelle le temps est venu. Imaginez mes émotions mêlées quand j'ai découvert qu'avant que je puisse en avoir ma version publiée dans un livre à proprement parler ⁸, elle avait déjà fait l'objet d'une satire dans un roman, *Jeu de société* de David Lodge (trad. 1991). C'est apparemment un thème brûlant chez les déconstructionnistes :

Selon Robyn (ou, plus précisément, selon les auteurs qui ont influencé sa pensée sur ces questions), il n'y a rien de tel que le « Moi » sur lequel s'appuient le capitalisme et le roman classique – c'est-à-dire une âme ou essence finie, unique, qui constitue l'identité d'une personne ; il n'y a qu'une position de sujet dans une toile infinie de discours – les discours du pouvoir, du sexe, de la famille, de la science, de la religion, de la poésie, etc. Et par la même occasion, il n'y a rien de tel qu'un auteur, c'est-à-dire quelqu'un qui est à l'origine d'une œuvre de fiction *ab nihilo*... selon les mots fameux de Jacques Derrida... « Il n'y a pas de hors texte. » Il n'y a pas d'origines, il n'y a que de la production, et nous produisons nos « moi » dans le langage. Non pas que « vous soyez ce que vous mangez », mais « vous êtes ce que vous parlez », ou, plutôt, « vous êtes ce qui vous parle », voilà la base axiomatique de la philosophie de Robyn, qu'elle appellerait, si elle devait lui donner un nom, du « matérialisme sémiotique ».

Du matérialisme sémiotique ? Dois-je l'appeler ainsi ? En dehors des allusions au capitalisme et au roman classique, sur lesquelles je ne me suis pas prononcé, ce passage facétieux est une jolie parodie de la thèse que je vais présenter. (Comme toute parodie, elle exagère ; je ne dirais pas qu'il n'y ait *rien* en dehors du texte. Il y a, par exemple, toutes les étagères de livres, les bâtiments, les corps, les bactéries...)

Robyn et moi pensons de la même façon – et bien sûr nous sommes *tous deux*, selon les analyses qui nous sont propres, des personnages de fiction d'un certain genre, bien que d'un genre légèrement différent.

NOTES

1. On peut trouver des variations sur ces thèmes dans Humphrey (1976, 1983) et dans Thompson, Palacios et Varela (à paraître).

2. Les philosophes sont couramment friands du concept d'*espèces naturelles*, réintroduit en philosophie par Quine (1969), qui peut à présent regretter d'être devenu une doublure pour le concept secrètement populaire d'*essences*. « Les

choses vertes, ou du moins les émeraudes vertes, sont une espèce », observe Quine (p. 116), qui manifeste sa propre appréciation du fait que, tandis que les émeraudes peuvent être une espèce naturelle, les choses *vertes* ne le peuvent probablement pas. La discussion présente a pour but de prévenir l'une des erreurs tentantes du naturalisme en fauteuil : le présupposé selon lequel, quoi que fasse la nature, c'est une espèce naturelle. Les couleurs ne sont *pas* des « espèces naturelles » précisément parce qu'elles sont le produit de l'évolution biologique, qui a une tolérance pour les frontières molles lorsqu'elle fait des catégories qui horrifieraient le philosophe qui cherche obstinément de bonnes et claires définitions. Si la vie d'un être quelconque dépendait de la mise dans le même sac de la lune, du fromage et des bicyclettes, vous pouvez être sûrs que Mère Nature lui trouverait le moyen de « voir » ceux-ci comme étant « intuitivement la même chose exactement ».

3. La primatologue Sue Savage-Rumbaugh m'a informé que les bonobos élevés en laboratoire ou les singes pygmées ne montrent aucun signe inné de dégoût pour les serpents, à l'inverse des chimpanzés.

4. La soudaineté serait importante, puisque si cela arrivait très progressivement, vous pourriez fort bien ne pas le remarquer. Comme Hardin (1990) l'a fait observer, le jaunissement progressif de vos lentilles avec l'âge modifie lentement votre sens des couleurs primaires ; si on vous montre un exemple des couleurs et qu'on vous demande d'indiquer un rouge pur (un rouge ne contenant aucun orange ni aucun pourpre) l'endroit du continu que vous indiquez est en partie fonction de votre âge.

5. « Le fait même que nous aimerions tellement dire : “ *Voici* la chose importante ” – alors que nous indiquons en privé la sensation – suffit à montrer combien nous avons tendance à dire quelque chose qui ne donne aucune information » Wittgenstein, 1953, i298.

6. Ce serait un acte de malhonnêteté intellectuelle désespérée de citer cette assertion en la sortant de son contexte !

7. Cf. Lockwood, 1989 : « Comment se serait sentie la conscience si elle s'était sentie comme des milliards de minuscules atomes se balançant chez elle ? » (p. 15-16).

8. J'ai présenté les idées principales dans mes réflexions sur Borgès dans *Vues de l'esprit* (Hofstadter et Dennett, 1981, p. 348-352) et les ai réunies dans un exposé, « The Self as the Center of Narrative Gravity », présenté au Colloque de Houston en 1983. En attendant la parution de ce volume du colloque, j'ai publié une version un peu raccourcie de mon exposé dans le *Times Literary Supplement* du 16-22 sept. 1988 sous le titre ennuyeux – qui n'était pas le mien – « Pourquoi tout le monde est un romancier ». La version originale, sous le titre *The Self as the Center of Narrative Gravity*, est toujours à paraître dans F. Kessel, P. Cole et D. Johnson, éd., *Self and Consciousness : Multiple Perspectives*, Hillsdale, NJ : Erlbaum.

Chapitre 12

La réalité des moi

« En feignant qu'il y ait une Machine, dont la structure fasse penser, sentir, avoir une perception, on pourra la concevoir agrandie en conservant les mêmes proportions, en sorte qu'on y puisse entrer comme dans un moulin. Et cela posé, on ne trouvera, en la visitant au-dedans, que des pièces qui poussent les unes les autres, et jamais de quoi expliquer une perception. »

Gottfried Wilhelm Leibniz (1646-1716),
Monadologie, § 17, édition E. Boutroux,
Paris, Delagrave, 1970, p. 149-150.

« Pour ma part, quand je pénètre le plus intimement dans ce que j'appelle *moi*, je bute toujours sur une perception particulière ou sur une autre, de chaud ou de froid, de lumière ou d'ombre, d'amour ou de haine, de douleur ou de plaisir. Je ne peux jamais me saisir, *moi*, en aucun moment sans une perception et je ne peux rien observer que la perception... Si quelqu'un pense, après une réflexion sérieuse et impartiale, qu'il a, de *lui-même*, une connaissance différente, il me faut l'avouer, je ne peux raisonner plus longtemps avec lui. Tout ce que je peux lui accorder, c'est qu'il peut être dans le vrai aussi bien que moi et que nous différons essentiellement sur ce point. Peut-être peut-il percevoir quelque chose de simple et de continu qu'il appelle *lui* : et pourtant je suis sûr qu'il n'y a pas en moi de pareil principe. »

David Hume, 1739, *Traité de la nature humaine*,
part. IV, section VI,

Paris, Aubier Montaigne, 1973, tome I, p. 343-344
(trad. fr. par A. Leroy).

Depuis l'aube de la science moderne au XVII^e siècle, un accord quasi unanime a régné sur le fait que le moi, quoi qu'il puisse être, serait invisible sous un microscope, et invisible aussi à l'introspection. Pour certains, cela semblait vouloir dire que le moi est une âme non physique, un fantôme dans la machine. Pour d'autres, que le moi n'est rien du tout, mais une fiction de l'imagination en proie à la fièvre métaphysique. Et pour d'autres encore, tout bonnement qu'un moi est d'une manière ou d'une autre une sorte d'abstraction, quelque chose dont l'existence n'est pas en quoi que ce soit mise en doute par son invisibilité. Après tout, pourrait-on dire, un centre de gravité est tout aussi invisible – et tout aussi réel. Est-ce que c'est bien assez réel ?

La question de savoir s'il existe vraiment des moi peut paraître susceptible de recevoir une réponse ridiculement simple, dans l'une ou l'autre direction : existons-nous ? Bien sûr ! La question présuppose sa propre réponse. (Après tout, qui est ce *je* qui a recherché en vain un moi, selon Hume ?) Y a-t-il des entités, soit *dans* votre cerveau, soit *en plus de* votre cerveau, qui contrôlent votre corps, pensent vos pensées, prennent vos décisions ? Bien sûr que non ! Une telle idée est soit une idiotie empirique (le « neurone pontifical » de James) soit un bobard métaphysique (le « fantôme dans la machine » de Ryle). Quand une question simple reçoit deux réponses, « Évidemment oui ! » et « Évidemment non ! », cela vaut la peine de considérer une voie moyenne (Dennett, 1991), même si elle est vouée à paraître à toutes les parties aller contre nos intuitions au départ – tout le monde est d'accord pour dire qu'une telle voie récuse tel ou tel fait évident !

Comment les êtres humains tissent un moi

« En plus, ils avaient l'air de passer beaucoup de temps à manger et à boire et à aller à des soirées, et Frensic, dont l'apparence avait tendance à se limiter au plaisir sensuel qu'il avait de mettre les choses en lui plutôt que dans autrui, avait quelque chose d'un gourmet. »

Tom Sharpe, 1977.

Le romancier Tom Sharpe suggère, dans ce passage drôle mais gênant, que lorsque vous allez vraiment au fond des choses, tout

plaisir sensuel consiste à jouer autour de sa propre frontière, ou de celle de quelqu'un d'autre. Il y a là quelque chose de juste. Ce n'est pas toute la vérité, mais c'est au moins une partie de la vérité.

Les gens ont un moi. Et les chiens ? Et les homards ? Si les moi sont quelque chose, alors ils existent. Or il y a des moi. Il fut un temps, il y a de cela des milliers (ou des millions, ou des milliards) d'années, où il n'y en avait aucun – du moins pas sur cette planète. Il faut donc bien qu'il y ait – simple question de logique – une histoire vraie à raconter sur *la manière dont il s'est mis à y avoir* des êtres dotés de moi. Cette histoire devra raconter – simple question de logique – le processus (ou la série de processus) mettant en jeu les activités ou les comportements de choses qui n'ont pas encore de moi – ou qui ne sont pas encore des moi – mais qui finalement seront à l'origine d'un nouveau produit, d'êtres qui sont, ou ont un moi.

Au chapitre 6, nous avons vu comment l'apparition de raisons était aussi l'apparition de frontières, la frontière entre « moi » et « le reste du monde », distinction que doit faire même la plus humble des amibes, à sa manière aveugle et ignorante. Cette tendance minimale à se distinguer d'autrui afin de se protéger est le moi biologique ; or même un moi aussi simple n'est pas une chose concrète mais une abstraction, un principe d'organisation. En outre, les frontières d'un moi biologique sont poreuses et indéfinies – autre exemple de la tolérance de Dame Nature à l'erreur, si le coût est justifié.

À l'intérieur des murs du corps humain on trouve moult et moult intrus, allant des bactéries et des virus en passant par les mites microscopiques qui vivent comme des troglodytes dans la niche écologique de notre peau et de notre scalp, jusqu'à de plus grands parasites – ces horribles ténias, par exemple. Ces intrus sont tous de minuscules et authentiques protecteurs de soi, mais certains d'entre eux, tels que les bactéries qui peuplent notre système digestif et sans lesquelles nous mourrions, sont des partenaires tout aussi essentiels à notre quête de l'autopréservation que les anticorps de notre système immunitaire. (Si la théorie de la biologiste Lynn Margulis (1970) est correcte, les mitochondries qui font le travail dans presque toutes les cellules de notre corps sont les descendants de bactéries avec lesquelles « nous » avons joint nos forces, il y a des milliards d'années de cela.) D'autres intrus sont des parasites tolérés – qu'il ne vaut

pas la peine d'éviter, apparemment – et d'autres encore sont vraiment des ennemis de l'intérieur, mortels si on ne les déloge pas et si on ne les extrait pas de soi.

Ce principe biologique fondamental de distinction du moi d'avec le monde, de l'intérieur d'avec l'extérieur produit certains échos remarquables dans les profondeurs les plus hautes de notre psychologie. Les psychologues Paul Rozin et April Fallon (1987) ont montré dans une série fascinante d'expériences sur la nature du *dégoût* qu'il y a une vague de fond puissante et méconnue de résistance aveugle à certains actes qui, rationnellement considérés, ne nous troublent pas. Par exemple, pourriez-vous, s'il vous plaît, avaler votre salive en ce moment précis ? Cet acte ne vous remplit pas de révolusion. Mais supposons que je vous demande d'aller chercher un verre propre et de cracher dans le verre, *puis* d'avalier la salive du verre. Dégoûtant ! Mais pourquoi ? Cela semble être dû à votre perception qu'une fois que quelque chose est en dehors de votre corps il ne fait plus partie de vous – il devient étranger et suspect – il a renoncé à sa citoyenneté et devient quelque chose à rejeter.

Les débordements sont donc soit des moments d'anxiété, soit, comme l'a indiqué Sharpe, quelque chose qu'il faut tout spécialement apprécier. De nombreuses espèces ont développé de remarquables constructions pour étendre leurs frontières territoriales, soit pour rendre le mauvais genre de débordement plus difficile, soit pour faciliter les bons. Les castors construisent des barrages, et les araignées tissent des toiles, par exemple. Quand l'araignée tisse sa toile, elle n'a pas à comprendre ce qu'elle est en train de faire ; Dame Nature a simplement doté son minuscule cerveau des routines nécessaires à l'effectuation de cette tâche biologiquement essentielle d'ingénieur. Les expériences sur les castors montrent que même leurs pratiques magnifiquement efficaces d'ingénieur sont au moins largement le produit de pulsions et de tendances innées qu'ils n'ont pas besoin de comprendre pour en profiter (les castors apprennent, et peuvent même enseigner les autres, mais pour l'essentiel ils sont mus par de puissants mécanismes innés qui contrôlent ce que le behavioriste B.F. Skinner appelait le renforcement négatif. Un castor se mettra furieusement en quête de quelque chose – n'importe quoi – pour arrêter le bruit de l'eau courante, et dans une expérience, on voit un castor se calmer parce qu'il fait un cataplasme de boue

tout autour du haut-parleur d'où sortait le gargouillement enregistré !

Le castor protège sa frontière extérieure avec des brindilles et de la boue et l'une de ses frontières internes avec de la fourrure. L'escargot rassemble du calcium dans sa nourriture et s'en sert pour exsuder une coquille dure ; le bernard-l'hermite trouve sa coquille de calcium toute prête, en prenant la coquille abandonnée d'un autre être, évitant élégamment le processus d'ingestion et d'exsudation. La différence n'est pas fondamentale, selon Richard Dawkins, qui indique que le résultat, dans chaque cas, qu'il appelle *le phénotype étendu* (1982), fait partie de l'équipement biologique fondamental des individus qui sont soumis aux forces sélectives qui dirigent l'évolution.

La définition d'un phénotype étendu ne s'étend pas seulement au-delà de la frontière « naturelle » des individus pour inclure l'équipement externe tel que les coquilles (et les équipements internes tels que les bactéries résidentes) ; il inclut souvent d'autres individus de la même espèce. Les castors ne peuvent s'en tirer tout seuls ; ils ont besoin du travail de l'équipe pour construire un seul barrage. Les termites doivent se grouper par millions pour construire leurs châteaux.

Et considérons les constructions architecturales époustouflantes de l'oiseau des tonnelles (*bower-bird*) australien (Borgia, 1986). Le mâle construit des tonnelles élaborées, des basiliques (mausolées, tombeaux, châsses) de cour avec une immense nef centrale, richement décorée d'objets brillants et colorés – avec pour couleur prédominante le bleu profond mais aussi des capsules de bouteilles, des morceaux de verre de couleur et autres artefacts humains – qui ont été ramenés de fort loin, réunis, et soigneusement arrangés au mieux dans la tonnelle (maison de verdure) pour impressionner la femelle à qui il fait la cour. L'oiseau des tonnelles, comme l'araignée, n'a pas vraiment besoin de comprendre ce qu'il fait ; il se trouve simplement à l'ouvrage, sans savoir pourquoi, créant un édifice qui est crucial pour son succès en tant qu'oiseau des tonnelles.

Mais les constructions les plus étranges et les plus merveilleuses de tout le monde animal sont les constructions extraordinaires et compliquées du primate, de l'*Homo Sapiens*. Tout individu normal de cette espèce construit un *moi*. De son cerveau, il tisse une toile de mots et de faits, et comme les autres êtres, il n'a pas à savoir ce qu'il fait ; il le fait, tout simplement. Cette toile

le protège, tout comme la coquille de l'escargot, lui procure des moyens de subsistance, tout comme la toile de l'araignée, et favorise ses projets sexuels, tout comme la tonnelle de l'oiseau des tonnelles. À la différence de l'araignée, un individu humain ne se contente pas d'*exsuder* sa toile ; sur le modèle plutôt du castor, il travaille dur pour réunir les matériaux à partir desquels il construira sa forteresse protectrice. Comme un oiseau des tonnelles, il s'approprie maints objets trouvés qui se trouvent le ravir – lui ou son compagnon –, y compris maints objets qui ont été conçus par d'autres pour d'autres fins.

Cette « toile de discours », comme l'appelait Robyn à la fin du chapitre précédent, est autant un produit biologique que n'importe laquelle des constructions que l'on peut trouver dans le monde animal. Si on le lui enlève, un individu humain est aussi incomplet qu'un oiseau sans ses plumes, qu'une tortue sans sa carapace. (Les vêtements, aussi, font partie du phénotype étendu de l'*Homo Sapiens* dans presque toute niche habitée par cette espèce. Une encyclopédie illustrée de zoologie ne devrait pas plus dépeindre l'*Homo Sapiens* nu qu'elle ne devrait dépeindre l'*Ursus arctus* – l'ours brun – en train de porter un habit de clown ou de conduire une bicyclette.)

L'organisation d'un termitier est si merveilleuse qu'il a semblé à certains observateurs que toute colonie de termites n'est que le résultat d'un million de petits agents semi-indépendants, chacun d'entre eux étant lui-même un automate, accomplissant ce qu'il a à faire. L'organisation d'un moi humain est si merveilleuse qu'il a semblé à bien des observateurs que tout être humain a également une âme : un Dictateur bienveillant dirigeant tout de son quartier général.

Dans toute ruche d'abeilles ou colonie de termites, il y a, à coup sûr, une reine des abeilles ou une reine des termites, mais ces individus sont plus patients que l'agent, ressemblent plus à ces joyaux de la couronne qu'il faut protéger qu'au chef des forces de protection – en fait leur nom royal est plus adapté aujourd'hui que dans des âges plus reculés, car ils ressemblent plus à la reine Elisabeth II qu'à la reine Elisabeth I^{re}. Il n'y a pas d'abeille Margaret Thatcher, pas de termite George Bush, pas de Bureau Oval dans la fourmilière.

Est-ce que nos moi, nos moi non minimalement *égoïques*, manifestent la même perméabilité et flexibilité de frontières que les moi plus simples d'autres créatures ? Est-ce que nous étendons

nos frontières personnelles – les frontières de nos *moi* – de manière à englober tout ce qui peut relever de nos « matériaux » ? En général, peut-être pas, mais il y a certainement des moments où cela paraît vrai, psychologiquement. Par exemple, si certains se contentent de posséder une voiture et de la conduire, d'autres sont des *automobilistes* ; l'automobiliste invétéré préfère *être* un agent consommateur d'essence à quatre roues plutôt qu'un agent consommateur de nourriture sur deux jambes, et son usage du pronom personnel à la première personne trahit cette identification :

Je ne fais pas bien mes virages les jours de pluie, parce que mes pneus commencent à être lisses.

Aussi élargissons-nous parfois nos frontières ; à d'autres moments, en réponse à des objections réelles ou imaginaires que nous percevons, nous laissons nos limites se rétrécir :

Je n'ai pas fait cela ! Ce n'était pas mon vrai moi qui parlais. Oui, les mots sont sortis de mes lèvres, mais je refuse de les reconnaître comme étant les miens.

Je vous ai rappelé ces discours familiers pour que vous établissiez des ressemblances entre nos moi et ceux des fourmis et des bernard-l'hermite, mais les discours attirent aussi l'attention sur la différence la plus importante : les fourmis et les bernard-l'hermite ne parlent pas. Le bernard-l'hermite est conçu de telle sorte qu'il veille à se procurer une coquille. Son organisation, pourrions-nous dire, *implique* une coquille, et à partir de là, en un sens *très* faible, *représente* tacitement le bernard comme ayant une coquille mais le bernard ne *se représente* pas en un sens plus fort comme ayant une coquille. Il ne s'occupe absolument pas de représentation de soi. À qui se représenterait-il ainsi lui-même et pourquoi ? Il n'a pas besoin de se souvenir de cet aspect de sa nature, puisque son organisation structurelle originelle prend soin de ce problème, et qu'il n'y a pas d'autres parties intéressées en vue. Quant aux fourmis et aux termites, comme nous l'avons noté, ils accomplissent leurs projets communautaires sans s'appuyer sur les moindres empreintes ou sur des édits qui leur auraient été explicitement communiqués.

Nous, en revanche, sommes constamment en train de nous

présenter à autrui, et à nous-mêmes, et donc de nous *représenter* – en langage et en geste – sur le mode externe et interne. La différence la plus évidente dans notre environnement qui expliquerait cette différence de notre comportement est le comportement lui-même. Notre environnement humain ne contient pas seulement de la nourriture et un abri, des animaux à combattre ou à fuir, et des congénères avec lesquels nous devons nous unir, mais des mots, des mots, encore des mots. Ces mots sont de puissants éléments de notre environnement que nous incorporons, ingérons, et évacuons aisément, en les tissant comme des toiles d'araignées en des chaînes autoprotectrices de *narrativité*. En vérité, comme nous l'avons vu au chapitre 6, quand nous introduisons ces mots, ces véhicules de mêmes, ils ont tendance à prendre le dessus, pour *nous* créer à partir des matériaux bruts qu'ils trouvent dans notre cerveau.

Notre tactique fondamentale d'autoprotection, d'autocontrôle et d'autodéfinition ne consiste pas à tisser des toiles ou à construire des barrages, mais à raconter des histoires, et plus particulièrement, à concocter et à contrôler l'histoire que nous racontons aux autres – et à nous-mêmes – sur ce que nous sommes. Et tout comme les araignées n'ont pas à penser, consciemment et délibérément, à la manière de tisser leur toile, et tout comme les castors, à l'inverse des ingénieurs humains professionnels, ne tracent pas consciemment et délibérément le plan des structures qu'ils construisent, nous (à l'inverse des conteurs d'histoire humains *professionnels*) n'envisageons pas consciemment et délibérément quelles narrations raconter et comment les raconter. Nos récits sont tissés, mais pour la plus grande part nous ne les tissons pas ; ils nous tissent. Notre conscience humaine, et notre ipséité narrative, en sont le produit, non la source.

Ces chaînes ou flux de narrativité jaillissent *comme* d'une seule et unique source – pas simplement au sens physique évident où elles s'écoulent d'une seule bouche, ou d'un seul crayon ou d'une seule plume, mais en un sens plus subtil : leur effet sur n'importe quel public est de les encourager (ou du moins d'essayer) à poser un agent unifié dont ils sont les mots, qu'ils concernent : bref, à poser un *centre de gravité narrative*. Les physiciens apprécient l'énorme simplification à laquelle vous parvenez quand vous posez un centre de gravité pour un objet, un seul et unique point relativement à quoi on peut calculer toutes les forces gravitationnelles. Nous autres (hétérophénoménologues) apprécions

l'énorme simplification à laquelle on parvient en posant un centre de gravité narrative pour un corps humain qui tisse de la narrativité. Comme le moi biologique, ce moi psychologique ou narratif est cependant une autre abstraction, non une chose du cerveau, mais c'est un aimant remarquablement robuste et presque tangible de propriétés, le « propriétaire des registres » de tout ce qui peut traîner çà et là sans être réclamé. Qui possède une voiture ? Vous. Qui possède vos vêtements ? C'est vous. Qui donc possède votre corps ? C'est vous ! Quand vous dites :

Ceci est *mon* corps

on ne considère certainement pas que vous dites :

Ce corps se possède lui-même.

Mais que pouvez-vous alors être en train de dire ? Si ce que vous dites n'est ni une tautologie bizarre et inutile (ce corps est son propre propriétaire, ou quelque chose de ce genre) ni que vous êtes une âme immatérielle ou un marionnettiste fantomatique qui possède et manipule ce corps à la manière dont vous possédez et manipulez votre voiture, que pourriez-vous bien vouloir dire d'autre ?

Combien de moi par client ?

Je pense que nous verrions plus clairement ce que

Ceci est mon corps

veut dire, si nous pouvions répondre à la question : « par opposition à quoi ? par opposition par exemple à ceci ? »

Non ; c'est *le mien* et je n'aime pas le partager.

Si nous pouvions voir quel effet cela ferait à deux moi (ou plus) de se battre pour avoir le contrôle d'un seul et unique corps, nous verrions mieux en quoi consiste vraiment un moi unique. En tant que savants du moi, nous aimerions nous livrer à des expériences contrôlées, dans lesquelles, en faisant varier

les conditions initiales, nous pourrions voir tout simplement ce qui doit arriver, dans quel ordre et exigeant quelles ressources, pour qu'un tel moi parlant puisse émerger. Y a-t-il des conditions dans lesquelles la vie continue et où aucun moi n'émerge ? Y a-t-il des conditions dans lesquelles plus d'un moi émerge ? Nous ne pouvons nous livrer pour des raisons éthiques à ce genre d'expériences, mais, comme si souvent auparavant, nous pouvons nous servir des données obtenues grâce aux expériences terribles auxquelles se livre la nature, en tirant prudemment des conclusions.

Une telle expérience est celle du Désordre de la Personnalité Multiple (MDP, *Multiple Personality Disorder*), dans lequel un seul corps humain a l'air d'être partagé en plusieurs moi, dont chacun a, de façon caractéristique, un nom propre et une autobiographie. L'idée du MDP frappe beaucoup de gens comme étant trop dépaysante et métaphysiquement bizarre pour qu'on puisse y croire – un phénomène paranormal à rejeter en même temps que les ESP, les rencontres rapprochées du troisième type et autres sorcières chevauchant des balais. Je soupçonne que certains de ces gens ont commis une erreur arithmétique simple : ils n'ont pas remarqué que deux ou trois ou dix-sept moi par corps n'a rien de plus métaphysiquement extravagant qu'un moi par corps. Il est déjà assez triste d'en avoir un !

« Je viens de voir passer une voiture dans laquelle il y avait cinq moi !

– Quoi ? L'esprit est pris de vertige ! De quel genre d'absurdité métaphysique s'agit-il ici ?

– Eh bien, il y avait aussi cinq corps dans la voiture.

– Ah, très bien, que ne l'avez-vous dit plus tôt ? Alors tout va bien.

– Ou peut-être seulement quatre corps, ou trois – mais à coup sûr cinq moi.

– Quoi ? ? ! !

L'arrangement normal est un moi par corps, mais si un corps peut en avoir un, pourquoi ne pourrait-il en avoir plus d'un dans des conditions anormales ?

Je ne veux pas suggérer par là qu'il n'y ait rien de choquant ou de profondément embarrassant à propos du MDP. C'est en fait un phénomène d'une étrangeté incomparable, non pas, à mon

avis, parce qu'il met au défi nos présupposés sur ce qui est *métaphysiquement* possible, mais plus parce qu'il met au défi nos présupposés sur ce qui est *humainement* possible, sur les limites de la cruauté et de la dépravation humaines d'un côté, sur les limites de la créativité humaine de l'autre. Car désormais il y a surabondance de preuves que ce n'est pas une poignée ou une centaine mais des milliers de cas de MDP qui sont diagnostiqués chaque jour, et cela est presque invariablement dû à des abus précoces sur l'enfant, le plus souvent sexuels, et d'une gravité écœurante. (Nicholas Humphrey et moi avons fait des recherches sur le MDP il y a plusieurs années de cela, et avons découvert qu'il s'agit d'un phénomène complexe qui s'étend bien au-delà du cerveau individuel celui qui en souffre.)

Ces enfants sont souvent maintenus dans des situations si extraordinairement terrifiantes et bouleversantes que je suis plus étonné qu'ils survivent tout simplement psychologiquement, que je ne le suis qu'ils réussissent à se préserver en dessinant, en un geste désespéré, un nouveau tracé de leurs frontières. Ce qu'ils font, lorsqu'ils sont confrontés à des souffrances et des conflits accablants, c'est ceci : ils « abandonnent ». Ils créent une frontière telle que l'horreur ne leur arrive pas à *eux* ; ou bien cela n'arrive à personne, ou bien cela arrive à quelqu'un d'autre, mieux à même de maintenir son organisation face à un tel assaut – du moins est-ce ce qu'ils *disent* avoir fait, pour autant qu'ils s'en souviennent.

Comment cela peut-il être ? Quel genre d'analyse donnerions-nous, finalement, au niveau biologique, d'un tel processus d'éclatement ? Faut-il qu'il y ait eu un seul et unique moi tout entier qui s'est en quelque sorte fissionné, comme une amibe ? Comment cela se pourrait-il si un moi n'est pas une partie physique à proprement parler d'un organisme ou d'un cerveau, mais, comme je l'ai suggéré, une abstraction ? La réponse au traumatisme paraît être si créative, en outre, que l'on est d'abord enclin à supposer que ce doit être l'œuvre de quelqu'un qui a tout supervisé ici : un programme cérébral de supervision, un contrôleur central, ou quelque chose de ce genre. Mais nous devons nous souvenir de la colonie de termites, qui paraissait aussi, au départ, exiger un chef exécutant central pour accomplir de tels projets intelligents.

Nous nous sommes habitués à des narrations évolutionnistes qui commencent par un état dans lequel un certain phénomène

n'existe pas encore et finissent par un état dans lequel le phénomène est décidément présent. L'innovation de l'agriculture, de l'habillement de l'habitat et des outils, l'innovation du langage, l'innovation de la conscience elle-même, l'innovation plus ancienne de la vie sur terre. Toutes ces histoires sont là pour être racontées. Et chacune d'entre elles doit traverser ce que nous pourrions appeler le gouffre de l'absolutisme. Ce gouffre est illustré par le curieux argument suivant (emprunté à Sanford, 1975) :

Tout mammifère a un mammifère pour mère.

Mais il n'y a eu qu'un nombre fini de mammifères, et donc il doit y avoir eu un premier mammifère,

Ce qui contredit notre première prémisse, et donc, contrairement aux apparences,

Il n'y a pas de choses telles que des mammifères !

Il faut laisser tomber quelque chose. Mais quoi ? Le philosophe absolutiste ou essentialiste est attiré par les lignes précises, les seuils, les « essences » et les « critères ». Pour l'absolutiste, il doit en vérité y avoir eu un premier mammifère, une première chose vivante, un premier moment de conscience, un premier agent moral ; qui fut n'importe quel produit d'un saut, n'importe quel nouveau candidat radicalement nouveau satisfaisant aux conditions essentielles – quelle que soit leur nature, révélée par l'analyse.

C'est ce goût pour les frontières entre les espèces qui a été le plus grand obstacle intellectuel que Darwin a dû affronter lorsqu'il a essayé de développer la théorie de l'évolution (Richards, 1987). À l'opposé de cette façon de penser, on trouve l'anti-essentialisme qui se sent à l'aise avec les cas pénombraux et le manque de lignes de division strictes. Comme les moi et les esprits et même la conscience elle-même sont des produits biologiques (et non des éléments qu'on peut trouver dans la table périodique de la chimie), nous devrions nous attendre à ce que les transitions entre eux, et les phénomènes qui ne le sont pas, soient progressives, litigieuses, truquées. Cela ne veut pas dire que tout soit toujours en transition, toujours progressif ; les transitions, qui ont l'air progressives vues de près, ressemblent habituellement à des ponctuations abruptes entre des plateaux d'équilibre vues de plus loin (Eldredge et Gould, 1972, mais voir aussi Dawkins, 1982, p. 101-109).

L'importance de ce fait pour les théories philosophiques (et

pour les goûts préférés des philosophes) n'est pas assez largement reconnue. Il y a toujours eu – et il y aura toujours – quelques choses transitoires, des « chaînons manquants » des quasi-mammifères et autres du même genre qui défient toute définition, mais le fait est que *presque toutes* les choses réelles (par opposition à celles qui ne sont que possibles) dans la nature ont tendance à tomber dans des groupes de ressemblance séparés par d'immenses gouffres béants. Nous n'avons pas besoin d'« essences » ou de « critères » pour empêcher le sens de nos mots de dérapier partout ; nos mots resteront là où ils sont, bien solidement arrimés comme s'ils étaient retenus par la gravité au groupe de ressemblance le plus proche, même s'il y a eu – ou doit y avoir eu – un bref isthme qui l'a jadis attaché par une série d'étapes progressives à quelque groupe voisin. Cette idée s'applique indiscutablement à de nombreux sujets. Mais bien des gens qui sont tout à fait à l'aise pour adapter cette approche pragmatique à la nuit et au jour, au vivant et au non-vivant, au mammifère et au pré-mammifère, s'inquiètent lorsqu'on les invite à adopter la même attitude envers le fait d'avoir un moi et de ne pas avoir un moi. Ils pensent que là, si c'est bien quelque part dans la nature, il doit y avoir Tout ou Rien et Un seul par client.

La théorie de la conscience que nous avons développée discrédite ces présupposés, et le Désordre de la Personnalité Multiple fournit une bonne illustration de la manière dont la théorie en constitue une objection. La conviction qu'il *ne peut* y avoir de quasi-moi ou de genres de moi, et qu'en outre il *doit* y avoir tout un nombre de moi associés à un corps – et il vaut mieux que ce soit le nombre un – ne va *pas* du tout de soi. C'est-à-dire qu'elle ne va pas du tout de soi, maintenant que nous avons développé de façon assez détaillée une autre voie possible que celle du Théâtre Cartésien avec son Témoin ou son Signifieur Central. Le MDP est bien d'un côté une objection à ces présupposés, mais nous pouvons aussi imaginer une objection qui viendrait de l'autre côté : deux corps ou plus partageant un seul et unique moi ! Le cas peut réellement se présenter, à York, Angleterre : les jumelles Chaplin, Greta et Freda (*Time*, 6 avril 1981). Ces vraies jumelles, qui ont aujourd'hui la quarantaine, et vivent ensemble à l'hôtel, paraissent agir *comme un seul et unique individu* ; elles collaborent en faisant par exemple des actes de parole uniques, finissant les phrases l'une de l'autre avec aisance ou parlant à l'unisson avec juste une fraction de seconde de décalage.

Pendant des années, elles ont été inséparables, aussi inséparables que pourraient le concevoir deux jumelles qui ne sont pas siamoises. Certains de ceux qui ont eu affaire à elles suggèrent que la tactique naturelle et efficace qui semble s'imposer est de les considérer plutôt comme un *elle*.

Notre thèse vient étayer la possibilité théorique non seulement du MDP mais aussi du FPD (*Fractional Personality Disorder*, Désordre de la Personnalité Fractionnelle). Est-ce possible ? Pourquoi pas ? Je ne suis pas un instant en train de suggérer que ces jumelles étaient liées par télépathie ou par ESP ou toute autre sorte de liens occultes. Je suggère qu'il y a une foule de moyens subtils et quotidiens de communication et de coordination (des techniques souvent hautement développées par des vrais jumeaux, en fait), et comme ces jumelles ont vu, entendu, touché, senti et pensé pas mal d'événements durant leur vie, et commencé, sans nul doute avec des cerveaux disposés assez semblablement, à réagir à ces stimuli, il pourrait ne pas y avoir besoin d'énormément de canaux de communication pour qu'elles se sentent chez elles sur la base d'une sorte de vague harmonie. (Et par ailleurs, quel degré d'unification atteint parmi nous celui qui possède au plus haut point un moi ?) Nous devrions hésiter à prescrire les limites d'une telle coordination effective.

Mais en tout cas, n'y aurait-il pas aussi deux moi clairement définis, l'un pour chaque jumelle, et qui explique qu'on soutienne cette curieuse charade ? Peut-être, mais que se passerait-il si chacune de ces femmes était devenue si peu égotiste (*so selfless*) (comme on dit) dans sa dévotion à la cause commune, qu'elle se soit plus ou moins perdue (*herself*) (comme on dit aussi) dans le projet ? Comme l'a dit un jour le poète Paul Valéry, en donnant un tour délicieux au mot de son compatriote : « Tantôt je suis, tantôt je pense ».

Au chapitre 10, nous avons vu que bien que la conscience semble continue, elle est en fait pleine de trous. Un moi pourrait être tout aussi troué, sombrant dans le néant aussi aisément que l'on souffle la flamme d'une bougie, pour se voir ranimé un peu plus tard, en des circonstances plus favorables. Êtes-vous cette personne même dont vous vous souvenez de façon imprécise des aventures au jardin d'enfants (tantôt avec vivacité, tantôt plus obscurément) ? Est-ce que les aventures de cet enfant, dont la trajectoire à travers l'espace et le temps a apparemment été en continuité avec celle de votre corps, sont bien vos propres aven-

tures ? Cet enfant qui a votre nom, un enfant dont la signature gribouillée sur un dessin au crayon vous rappelle la manière dont *vous* aviez l'habitude de signer votre nom – est-ce que cet enfant est (était) bien vous ? Le philosophe Derek Parfit (1984) a comparé une personne à un club, un genre plutôt différent de construction humaine, qui pourrait cesser telle année d'exister, et pourrait être reconstitué par certains de ses membres (antérieurs ?) quelques années plus tard. S'agirait-il du même club ? Peut-être, si, par exemple, le club avait eu une constitution écrite aménageant justement ce genre de ruptures d'existence. Mais il pourrait n'y avoir aucune manière de décider. Nous pourrions connaître tous les faits qui pourraient de façon concevable peser sur la situation et voir qu'ils ne permettent pas de conclure quant à l'*identité* du (nouveau ?) club. Du point de vue des moi – ou des personnes – qui apparaissent ici, c'est la bonne analogie ; les moi ne sont pas des perles de l'âme qui existent de façon indépendante, mais les artefacts des processus sociaux qui nous créent, et comme d'autres artefacts de ce genre, ils sont soumis à des changements brusques de statut. La seule « force » qui revienne à la trajectoire d'un moi, ou d'un club, c'est la stabilité qui lui est impartie par la toile de croyances qui le constitue, et quand ces croyances se rompent, elle se rompt, soit pour toujours, soit temporairement.

Il est important de garder cela à l'esprit lorsqu'on considère un autre sujet favori des philosophes, le phénomène abondamment discuté des patients qui ont subi une fission du cerveau. Une telle fission du cerveau est le résultat de la *commisurotomie*, opération qui sépare en deux le corps calleux, la large bande de fibres qui relie directement les hémisphères gauche et droit du cortex. Cela laisse les hémisphères toujours reliés de façon indirecte, à travers toute une série de structures du moyen cerveau, mais c'est à l'évidence une procédure terrible, à laquelle il ne faut se livrer que s'il n'y a plus rien d'autre à faire. Cela soulage le patient dans certains cas graves d'épilepsie que l'on ne peut pas traiter autrement, en empêchant les tempêtes électriques générées de l'intérieur qui causent les crises de dévaster le cortex en passant d'un « centre » d'origine situé dans un hémisphère au côté opposé. La légende philosophique standard veut que les patients qui ont connu une fission du cerveau peuvent être « fissionnés en deux moi » mais, à part cela, ne souffrent d'aucune sérieuse diminution de leurs possibilités au terme de l'opération chirurgicale. La ver-

sion la plus attirante de cette simplification extrême est que les deux « côtés » de la personne d'origine – l'hémisphère gauche, analytique, rigoureux, et l'hémisphère droit, en retrait, intuitif et holistique – sont après l'opération dispensés de rayonner d'une plus grande individualité, maintenant que le proche et normal travail d'équipe doit être remplacé par une *détente* moins intime. C'est une idée attirante, mais c'est une exagération outrancière des découvertes empiriques qui l'inspirent. En fait, ce n'est que dans une infime fraction de cas que l'on observe *l'un quelconque* des symptômes théoriquement frappants d'ipséité multiple (voir, par ex., Kinsbourne, 1974 ; Kinsbourne et Smith, 1974 ; Levy et Trevarthen, 1976 ; Gazzaniga et Ledoux, 1978 ; Gazzaniga, 1985 ; Oakley, 1985 ; Dennett, 1985b).

Il n'est pas surprenant que les patients au cerveau divisé, comme les patients aveugles et les gens qui ont un désordre de la personnalité multiple, ne soient pas à la hauteur de leur affiche philosophique et ce n'est la faute de personne. Ce n'est pas que les philosophes (et bien d'autres interprètes, y compris les principaux chercheurs) exagèrent délibérément leurs descriptions des phénomènes. C'est plutôt que dans leur effort pour décrire les phénomènes avec concision, ils trouvent que les ressources limitées du langage ordinaire les dirigent inexorablement vers le modèle simpliste du Chef du Corps, du Fantôme dans la Machine, de l'Auditoire du Théâtre Cartésien. Nicholas Humphrey et moi, en comparant prudemment nos propres notes sur ce qui s'était produit lors des diverses rencontres que nous avons eues avec des patients atteints de MDP, nous sommes aperçus que nous nous sommes souvent laissés aller, malgré nous, à des tournures de phrases si naturelles mais gravement fourvoyantes pour décrire ce que nous avons réellement vu. Thomas Nagel (1971), premier philosophe à écrire sur les patients au cerveau divisé, a présenté une analyse judicieuse et exacte des phénomènes tels qu'on les comprenait alors, et, notant la difficulté qu'il y avait à fournir une analyse cohérente, a fait la supposition suivante : « Il nous est peut-être impossible d'abandonner certaines façons de nous concevoir et de nous représenter, aussi faible que soit le soutien qu'ils puissent recevoir de la recherche scientifique. »

C'est en effet difficile mais pas impossible. Le pessimisme de Nagel est lui-même exagéré. N'avons-nous pas réussi, en fait, à nous débarrasser de la manière traditionnelle de penser ? Certes, il se peut que certains ne *veulent pas* abandonner la vision

traditionnelle. Il pourrait même y avoir de bonnes raisons – des raisons morales – d'essayer de préserver le mythe des moi, ces perles du cerveau, ces choses particulières, concrètes et dénombrables, au lieu d'abstractions, et de refuser d'approuver la possibilité de quasi-moi, de semi-moi, de moi-de-transition. Mais c'est sûrement la manière correcte de comprendre les phénomènes de fission cérébrale. Pendant de brèves périodes au cours des procédures expérimentales soigneusement conduites, quelques-uns des patients opèrent une bifurcation dans leur réponse à tel ou tel sujet, créant temporairement un second centre de gravité narrative. Quelques effets de la bifurcation peuvent s'attarder indéfiniment en des traces mémorielles mutuellement inaccessibles, mais à part ces traces vraiment tout à fait primitives de la bifurcation, la vie d'un second moi rudimentaire dure au plus quelques minutes, ce qui ne laisse pas beaucoup de temps pour se nantir de la sorte d'autobiographie dont sont constitués les moi au complet. (Cela est juste aussi évidemment vrai de la plupart de la douzaine de moi fragmentaires développés par des patients de MDP ; il n'y a tout simplement pas assez d'heures de veille dans la journée pour que la plupart d'entre eux puissent mettre de côté plus de quelques minutes de biographie exclusive par semaine. Voir l'appendice C.)

Le caractère distinct de différents récits est l'essence vitale de différents moi. Comme le note le philosophe Ronald de Sousa (1976) :

« Quand le Dr. Jekyll se transforme en Mr. Hyde, c'est là une chose étrange et mystérieuse. S'agit-il de deux personnes qui occupent chacun à son tour un seul corps ? Mais il y a quelque chose de plus étrange : le Dr. Juggle et le Dr. Boggle eux aussi occupent chacun à son tour un seul corps. *Mais ils sont aussi semblables que de vrais jumeaux !* Vous contre-attaquez : Eh bien, dites qu'ils se sont changés l'un en l'autre ? Après tout, pourquoi pas : si le Dr. Jekyll peut se changer en un homme aussi différent que Hyde, il doit assurément être d'autant *plus facile* pour Juggle de se changer en Boggle, qui est exactement comme lui.

Nous avons besoin du conflit ou de la différence forte pour ébranler l'hypothèse naturelle qu'à un corps correspond au plus un agent » (p. 219).

Quel effet cela fait-il donc d'être le Moi de l'Hémisphère Droit chez un patient au cerveau divisé ? C'est la question la plus naturelle du monde ¹, et qui permet d'éviter une image qui fait sursauter – et refroidit – l'esprit : vous voilà, pris au piège dans l'Hémisphère Droit d'un corps dont vous connaissez intimement

(et continuez de contrôler) le côté gauche et dont le côté droit est à présent aussi éloigné que le corps d'un étranger de passage. Vous aimeriez dire au monde quel effet vous fait d'être vous, mais rien à faire ! Vous êtes coupés de toute communication verbale par la perte de vos lignes téléphoniques indirectes avec la station de radio de l'Hémisphère Gauche. Vous faites de votre mieux pour signaler votre existence au monde extérieur, tirant violemment sur la moitié de votre visage pour faire des froncements de sourcil et des sourires de travers, et à l'occasion (si vous êtes un Moi virtuose de l'Hémisphère Droit) en griffonnant un mot ou deux avec votre main gauche.

Cet exercice d'imagination pourrait se poursuivre selon les manières évidentes, mais nous savons que c'est une fantaisie — tout autant une fantaisie que les histoires charmantes qu'on trouve dans Beatrix Potter sur Peter Rabbit et ses amis animaux anthropomorphiques. Non parce que « la conscience ne se trouve que dans l'hémisphère gauche », et non parce qu'*il ne se pourrait pas* que quelqu'un se soit trouvé dans d'aussi sales draps, mais simplement parce que *ce n'est pas vrai* que la commissurotomie laisse dans son sillon des organisations à la fois assez distinctes et robustes pour soutenir un tel moi séparé.

Ce pourrait difficilement être un argument contre ma théorie du moi qu'il soit « logiquement possible », qu'il y ait un tel Moi de l'hémisphère droit dans un patient au cerveau divisé, car ma théorie dit qu'il n'y en a pas. Elle dit aussi pourquoi : les conditions nécessaires à l'accumulation du genre de richesse narrative (et l'indépendance) qui constitue un moi « au complet » ne sont pas présentes. Ma théorie est de même imperméable à la thèse — que je ne songerai pas à nier — qu'il *pourrait* y avoir des Jeannot-lapins parlants, des araignées qui écrivent des messages en anglais dans leurs toiles, et à cet égard, des trains ou autres Tchou Tchous mélancoliques. Ce serait possible, je suppose, mais il n'y en a pas — aussi ma théorie n'a-t-elle pas à les expliquer.

L'insoutenable légèreté de l'être

« Tout ce qui arrive, ici ou là, nous sommes enclins à nous demander qui ou quoi en est responsable. Ce qui nous conduit à découvrir des explications que nous ne pourrions imaginer autrement, et ce qui nous aide à prédire et contrôler non seulement ce qui arrive dans le monde, mais aussi ce qui arrive à notre esprit. Mais qu'en serait-il si ces mêmes tendances devaient nous conduire à imaginer des choses et des causes qui n'existent pas ? Alors nous inventerions de faux dieux et des superstitions et y verrions la main dans toute coïncidence fortuite. En vérité, peut-être que ce mot étrange " Je " – tel qu'on l'emploie par exemple dans " Je viens d'avoir une bonne idée " – reflète la même tendance. Si vous êtes obligés de trouver une cause qui cause tout ce que vous faites – eh bien, alors, il faut un nom à ce quelque chose. Vous l'appellez "*moi*". Je l'appelle "*vous*". »

Marvin Minsky, 1986, p. 232.

Un moi, selon ma théorie, n'est pas un vieux point mathématique, c'est une abstraction définie par les myriades d'attributions et d'interprétation (y compris les attributions de soi et les interprétations de soi) qui ont composé la biographie du corps vivant dont il est le Centre de Gravité Narrative. Comme tel, il joue un rôle singulièrement important dans l'économie cognitive incessante de ce corps vivant, parce que, de toutes les choses qui se trouvent dans l'environnement et dont un corps actif doit faire des modèles mentaux, aucune n'est plus cruciale que le modèle que l'agent a de lui-même (voir par exemple, Johnson-Laird (1988 ; Perlis, 1991)).

Pour commencer, tout agent doit savoir ce qu'il est dans le monde ! Cela peut sembler à première vue trivial ou impossible. « Je suis moi ! » n'apporte pas vraiment d'information. Que pourrait-on avoir besoin de savoir d'autre – ou que pourrait-on découvrir si on ne le savait pas déjà ? Pour des organismes plus simples, c'est vrai, il n'y a réellement pas grand-chose en matière de connaissance de soi au-delà de la sagesse biologique rudimentaire religieusement conservée dans des maximes telles que Quand

vous avez faim, ne vous Mangez pas Vous-mêmes ! et Quand il y a de la Douleur, c'est la Vôtre ! Dans tout organisme, y compris les êtres humains, la reconnaissance de ces principes biologiques structurels de base est simplement « câblée » – elle fait partie de la structure sous-jacente du système nerveux, comme de cligner de l'œil lorsque quelque chose s'en approche, ou de frissonner lorsqu'il fait froid. Un homard pourrait bien manger les pinces d'un autre homard, mais l'idée qu'il pourrait manger l'une de ses propres pinces lui paraît heureusement impensable. Ses options sont limitées, et quand il « pense à » bouger une pince son « penseur » est câblé directement et de façon appropriée à la pince même qu'il pense bouger. Avec les êtres humains (et les chimpanzés et peut-être quelques autres espèces) en revanche, il y a plus d'options, et donc plus de sources de confusion.

Il y a quelques années, les autorités du port de New York ont fait l'expérience d'un système de radar commun pour les propriétaires de petits bateaux. Une unique antenne de radar basée à terre formait une image radar du port qui pouvait alors être transmise comme un signal de télévision aux propriétaires de bateaux qui pouvaient s'épargner le coût du radar en installant simplement des petits postes de télévision sur leur bateau. À quoi cela servait-il ? Si vous vous perdiez dans le brouillard, et que vous regardiez l'écran de télévision, vous saviez que l'un de ces nombreux points lumineux mobiles sur l'écran, c'était vous – mais lequel ? Voici un cas où la question « quelle chose suis-je dans le monde ? » n'appelle une réponse ni triviale ni impossible. Le mystère se ramène à un truc bien simple : faites rapidement tourner votre bateau en un cercle resserré ; alors votre point mobile est celui qui trace le petit « O » sur l'écran – à moins que plusieurs bateaux perdus dans le brouillard essaient de faire le même test au même moment.

La méthode n'est pas sûre à cent pour cent, mais elle marche la plupart du temps, et elle illustre bien un point encore plus général : pour contrôler le genre d'activités sophistiquées auxquelles se livrent les corps humains, le système de contrôle du corps (logé dans le cerveau) doit être capable de reconnaître qu'une grande variété d'entrées l'informent sur lui-même, et lorsque des dilemmes se présentent, ou que le scepticisme s'installe, la seule manière fiable (mais qui n'est pas sûre à cent pour cent) de s'en sortir et d'attribuer correctement cette information c'est d'effectuer de petites expériences : faire quelque chose et

chercher à voir ce qui bouge². Un chimpanzé peut aisément apprendre à atteindre des bananes à travers le trou du mur de sa cage, guidant les mouvements de son bras en regardant son propre bras sur un circuit fermé d'un téléviseur installé à une assez bonne distance de son bras (Menzel et al., 1985). C'est là décidément un exemple non trivial de reconnaissance de soi, puisqu'elle dépend de ce que l'on remarque la consonance entre les mouvements du bras qui sont vus sur l'écran et les mouvements invisibles mais *voulus* du bras. Qu'arriverait-il si les expérimentateurs introduisaient un léger retard dans l'enregistrement vidéo ? Combien de temps pensez-vous que cela mettrait pour que *vous* découvriez que vous étiez en train de regarder votre propre bras (sans indications verbales du dispositif expérimental) si un retard d'enregistrement de disons vingt secondes était introduit dans le circuit fermé ?

Le besoin de connaissance de soi s'étend au-delà des problèmes d'identification des signes extérieurs de notre propre mouvement corporel. Nous avons besoin de connaître nos propres états, tendances, décisions, forces et faiblesses internes, et la méthode de base pour parvenir à cette connaissance est essentiellement la même : faites quelque chose et « cherchez » à voir ce qui « bouge ». Un agent avancé doit échafauder des pratiques pour prendre note de ce qui peut lui arriver, dans son corps et « mentalement ». Chez les êtres humains, comme on l'a vu, ces pratiques impliquent que l'on ne cesse de raconter et de contrôler des histoires, dont certaines sont factuelles et d'autres fictionnelles. Les enfants pratiquent cela à haute voix (pensez à Snoopy, qui se dit du haut de sa niche : « Voici l'as de l'aviation de la première guerre mondiale »). Nous autres adultes le faisons plus élégamment : en silence, tacitement, sans le moindre effort, prenant note des différences entre nos fantaisies et nos reconstitutions et réflexions « sérieuses ». Le philosophe Kendall Walton (1973, 1978, 1981) et le psychologue Nicholas Humphrey (1986) ont montré à partir de différentes perspectives l'importance que revêtent les drames, les récits et le phénomène plus fondamental du faux-semblant dans l'exercice qu'ils apportent aux êtres humains qui sont des tisseurs de moi novices.

Ainsi échafaudons-nous une histoire définie sur nous-mêmes, organisée autour d'un point mobile lumineux fondamental de représentation de soi (Dennett, 1981). Le point n'est pas un moi, bien sûr ; c'est une *représentation* de moi (et le point sur l'écran

radar indiquant l'île Ellis n'est pas une île – c'est une représentation d'île). Ce qui fait d'un point le point-*moi* et d'un autre point simplement un point-*lui*, ou un point-*elle*, ou un point-*quelque chose*, n'est pas ce à quoi il ressemble, mais ce à quoi il sert. Il réunit et organise l'information sur le sujet *moi* de la même manière que d'autres structures dans mon cerveau prennent note de l'information sur Boston, ou Reagan, ou la crème glacée.

Et où se trouve la chose sur laquelle *porte* votre représentation de soi ? N'importe où où vous vous trouvez (Dennett, 1978b). Et qu'est-ce que cette chose ? Rien de plus et rien de moins que votre centre de gravité narrative.

Otto revient :

« Le problème avec les centres de gravité est qu'ils ne sont pas réels ; ce sont des fictions de théoricien.

– Ce n'est pas ce qui fait problème avec les centres de gravité ; c'est ce qui fait leur gloire. Ce sont des fictions *magnifiques*, des fictions que quiconque serait fier d'avoir créées. Et les personnages fictionnels de la littérature sont encore plus merveilleux. Pensez à Ismael, dans *Moby Dick*. “ Appelez-moi Ismael. ” C'est ainsi que s'ouvre le texte, et nous nous exécutons avec obligeance. Nous n'appelons pas le texte Ismael, ni Melville Ismael. Qui et quoi appelons-nous Ismael ? Nous appelons Ismael Ismael, le merveilleux personnage qu'on peut trouver dans les pages de *Moby Dick*. “ Appelez-moi Dan ”, entendez-vous de mes lèvres, et vous vous exécutez avec obligeance, non en appelant mes lèvres Dan, ou mon corps Dan, mais en *m'appelant* Dan, la fiction du théoricien créée par... eh bien, non par moi, mais par mon cerveau, agissant de concert avec mes parents, mes enfants et mes amis.

– C'est peut-être très bien pour vous. Mais *moi* je suis parfaitement réel. J'ai peut-être été créé par le processus social auquel vous venez de faire allusion (j'ai dû l'être, si je n'existais pas avant ma naissance), mais ce que le processus a créé est un moi *réel*, pas un simple personnage de fiction !

– Je crois savoir où vous voulez en venir. Si un moi n'est pas une chose réelle, qu'advient-il de la responsabilité morale ? C'est l'un des rôles les plus importants du moi dans notre schème conceptuel traditionnel et l'endroit où se fait la donne (*buck*), comme le disait le panneau d'Harry Truman. Si les moi ne sont pas réels – ne sont pas *réellement* réels – est-ce que la donne ne

continuera pas à passer de main en main, d'aller et venir, sans jamais s'arrêter ? S'il n'y a pas de Bureau Ovalé dans le cerveau, abritant une Autorité Suprême à laquelle on peut faire appel de toutes les décisions, la menace semble alors peser d'une bureaucratie kafkaïenne d'homoncules, qui répondent toujours, quand on leur oppose des arguments : " Ne me blâmez pas. Je ne fais que travailler ici. " La tâche de construction d'un moi qui peut *prendre* des responsabilités est un projet social et éducatif majeur, et vous avez raison de vous sentir concernés par les menaces qui pourraient peser sur son intégrité. Mais une perle cérébrale, un tout-ce-qu'on-voudra réel, " intrinsèquement responsable " est une babiole pathétique à brandir comme un gri-gri face à cette menace. Le seul espoir, et qui n'est pas vain, est de finir par comprendre, de façon naturaliste, la manière dont les cerveaux font naître des représentations de soi, équipant par là même les corps qu'ils contrôlent de moi responsables quand tout va bien. Le libre arbitre et la responsabilité morale valent bien la peine qu'on les recherche, et comme j'essaie de le montrer dans *Elbow Room : The Varieties of Free Will Worth Wanting* (1984), la meilleure défense qu'on puisse leur offrir, c'est d'abandonner le mythe désespérément pétri de contradictions de l'âme distincte et séparée.

— Mais est-ce que je n'existe pas ?

— Bien sûr que si. Vous voilà, assis sur cette chaise, en train de lire mon livre et de soulever des objections. Et assez curieusement, votre incarnation ordinaire, bien qu'étant une précondition nécessaire de votre création, n'est pas nécessairement requise pour le prolongement indéfini de votre existence. Mais si vous étiez une âme, une perle d'une substance immatérielle, nous ne pourrions " expliquer " votre immortalité potentielle qu'en postulant qu'il s'agit d'une propriété inexplicable, une *virtus dormitiva* inéliminable de matière animée. Et si vous étiez une perle de substance matérielle, un groupe spectaculairement spécifique d'atomes cérébraux, votre mortalité dépendrait des forces physiques qui les font tenir ensemble (nous pourrions demander aux physiciens en quoi consiste la " moitié de vie " d'un moi). Si vous pensez à vous comme à un centre de gravité narrative, en revanche, votre existence dépend de la persistance de cette narration (qui se rapproche plutôt des *Mille et Une Nuits*, mais en un seul et unique récit), qui pourrait *théoriquement* survivre indéfiniment à de nombreuses inversions de *medium*,

être téléportée aussi aisément (en principe) que le journal du soir, et indéfiniment stockée comme de la pure et simple information. Si ce que vous êtes, c'est une organisation de l'information qui a structuré le système de contrôle de votre corps (ou pour dire les choses sous leur forme provocatrice plus habituelle, si ce que vous êtes, c'est le programme qui se déroule dans votre ordinateur cérébral), alors vous pourriez en principe survivre à la mort de votre corps en restant aussi intact qu'un programme qui peut survivre à la destruction de l'ordinateur sur lequel il a été créé et s'est déroulé la première fois. Certains penseurs (par exemple Penrose, 1989) trouvent qu'il s'agit là d'une implication épouvantable et profondément contre-intuitive de la thèse que j'ai défendue ici. Mais si c'est d'une immortalité potentielle que vous avez la nostalgie, les autres voies possibles sont tout simplement indéfendables. »

NOTES

1. Il est intéressant de noter que Nagel, en 1971, s'occupait déjà explicitement de cette question (MQ, p. 160), avant de porter son attention sur les chauves-souris – sujet que nous discuterons au prochain chapitre.

2. Et comment savons-nous que *nous* sommes en train de faire quelque chose ? Où acquérons-nous cette petite dose initiale de connaissance de soi que nous utilisons pour cette action de levier ? Cela a paru être une question absolument fondamentale à certains philosophes (Castañeda, 1967, 1968 ; Lewis, 1979 ; Perry, 1979), et a été l'occasion d'une multitude d'écrits plus complexes les uns que les autres. S'il s'agit là d'un problème philosophique substantiel, il doit y avoir quelque chose de faux dans la réponse « triviale » (mais je ne vois pas quoi) : nous acquérons notre connaissance de soi fondamentale et originelle de la même façon que le homard ; nous sommes tout simplement câblés ainsi.

Chapitre 13

La conscience imaginée

Imaginer un robot conscient

Nous avons expliqué les phénomènes de la conscience humaine dans les chapitres précédents en termes d'opérations d'une « machine virtuelle », une sorte de programme d'ordinateur qui a évolué (et qui évolue) et qui façonne les activités du cerveau. Il n'y a pas de Théâtre Cartésien ; il n'y a que des Versions Multiples composées de processus de fixation de contenu qui jouent divers rôles semi-indépendants dans l'économie cérébrale plus vaste de contrôle du voyage du corps humain à travers son existence. La conviction étonnamment enracinée qu'il y a un Théâtre Cartésien est le résultat de diverses illusions cognitives que nous avons à présent exposées et remplacées par des états dispositionnels complexes du cerveau, et le moi (connu aussi sous les noms d'Auditoire dans le Théâtre Cartésien, de Signifieur Central, ou de Témoin) se révèle être une abstraction utile, une fiction théorique plutôt qu'un observateur ou un chef interne.

Si le moi n'est « que » le Centre de Gravité Narrative, et si tous les phénomènes de la conscience humaine sont explicables comme n'étant « que » les activités d'une machine virtuelle réalisée dans les connexions astronomiquement ajustables d'un cerveau humain, alors, en principe, un robot bien « programmé », avec un ordinateur composé d'éléments de silicone, serait conscient, aurait un moi. Ou plutôt il y aurait un moi conscient dont le corps serait le robot et dont le cerveau serait l'ordinateur. Certaines personnes trouvent évidente et incontestable cette conséquence de ma théorie. « *Bien sûr* que nous sommes des

machines ! Nous ne sommes que des machines très, très compliquées, évoluées, composées de molécules organiques plutôt que de métal et de silicone, et nous sommes conscients. Donc il y a des machines *conscientes* – à savoir nous. » Pour ces lecteurs cette conséquence était prévue d'avance. Ce qui s'est révélé intéressant pour eux, je l'espère, c'est la diversité des conséquences que nous avons rencontrées en cours de route, en particulier celles qui montrent à quel point l'image cartésienne de sens commun devait être remplacée au fur et à mesure que nous apprenions plus sur la machinerie effective du cerveau.

D'autres personnes, par contre, trouvent que la conséquence selon laquelle il pourrait, en principe, exister un robot conscient est tellement incroyable que cela revient à leurs yeux à une réduction à l'absurde de ma théorie. L'un de mes amis a un jour réagi à ma théorie avec candeur : « Mais, Dan, je ne peux tout simplement pas imaginer un robot conscient ! » Certains lecteurs peuvent être enclins à l'admettre. Ils devraient résister à cette inclination, parce qu'il s'était abusé. Son erreur était simple, mais révélatrice d'une confusion fondamentale qui interdit tout progrès dans la compréhension du phénomène de la conscience. « Tu *sais* que c'est faux lui ai-je répondu. Tu as souvent imaginé des robots conscients. L'important n'est pas que tu ne puisses pas imaginer un robot conscient ; c'est que tu ne puisses pas imaginer *comment* un robot pourrait être conscient ! »

Quiconque a vu R2D2 et C3PO dans *La Guerre des étoiles*, ou entendu Hal dans *2001*, a imaginé un robot conscient (ou un ordinateur conscient – que le système soit mobile, comme R2D2, ou immobile, comme Hal, ce n'est pas si crucial pour mettre en marche notre imagination). C'est littéralement un jeu d'enfant que d'imaginer le flux de la conscience d'un objet « inanimé ». Les enfants le font tout le temps. Non seulement les Nounours ont des vies intimes, mais aussi le Gentil Petit Train Électrique. Les sapins baumiers se tiennent debout silencieux dans les bois, craignant la hache du bûcheron, mais en même temps ils meurent d'envie de devenir des arbres de Noël dans quelque jolie maison bien chaude, entourés d'enfants heureux. La littérature pour enfants (pour ne rien dire de la télévision) est bourrée de possibilités d'imaginer les vies conscientes de pauvres choses de ce genre. Les artistes qui illustrent leurs histoires aident en général l'imagination des enfants en dessinant des visages expressifs sur ces drôles d'agents, mais ce n'est pas l'essentiel. Parler – comme

le fait Hal – suffira aussi bien, en l'absence d'un visage expressif, pour assurer l'illusion qu'il y a quelqu'un là-dedans, que cela fait un certain effet d'être Hal, ou un Nounours, ou un Tchou Tchou.

C'est là la difficulté, bien entendu : ce sont là des illusions – en apparence en tout cas. Il y a des différences entre tous ces objets. Il est évident qu'aucun Nounours n'est conscient, mais il n'est vraiment pas évident qu'un robot ne puisse pas l'être. Étant donné que mon ami trouvait difficile d'imaginer *comment* un robot pourrait être conscient, il éprouvait de la réticence à imaginer qu'un robot *soit* conscient – bien qu'il eût aisément pu le faire. Il y a toute la différence du monde entre ces deux sortes d'actes d'imagination, mais les gens tendent à les confondre. Il est bien entendu épouvantablement difficile d'imaginer comment le cerveau-ordinateur d'un robot pourrait être le support d'une conscience. Comment un amas compliqué d'événements de traitement d'information dans un tas de puces de silicone *pourrait-il* être équivalent à des expériences conscientes ? Mais il est tout aussi difficile d'imaginer la façon dont un cerveau humain organique pourrait servir de support à de la conscience. Comment un amas compliqué d'interactions électrochimiques entre des milliards de neurones *pourrait-il* être équivalent à des expériences conscientes ? Et pourtant nous imaginons aisément que des êtres humains soient conscients, même si nous ne pouvons toujours pas imaginer *comment* c'est possible.

Comment le cerveau pourrait-il être le siège de la conscience ? Les philosophes ont traité en général cette question comme une question rhétorique : ils ont suggéré que la réponse outrepasserait les pouvoirs de l'entendement humain. L'un des objectifs fondamentaux de ce livre a été de démolir ce préjugé. J'ai soutenu que vous *pouvez* imaginer comment cet amas compliqué d'activité dans le cerveau revenait à de l'expérience consciente. Mon argument est simple : je vous ai montré comment le faire. Il se trouve que la façon dont on peut imaginer cela consiste à penser le cerveau comme une certaine sorte d'ordinateur. Les concepts de l'informatique fournissent à notre imagination les béquilles dont nous avons besoin si nous trébuchons dans la *terra incognita* qui se trouve entre notre phénoménologie telle que nous la connaissons « par introspection » et nos cerveaux tels que la science nous les révèle. En considérant nos cerveaux comme des systèmes de traitement d'information, nous pouvons progressivement dissiper

le brouillard et frayer notre chemin à travers le grand fossé, en découvrant comment nos cerveaux peuvent produire tous ces phénomènes. Il y a beaucoup de chausse-trappes à éviter – des impasses attirantes comme le Signifieur Central, le « remplissement », et les « qualia », par exemple – et il ne fait pas de doute qu'il reste encore des confusions et des erreurs pures et simples dans l'ébauche que j'ai fournie, mais nous voyons du moins à quoi ressemblerait le chemin pour en sortir.

Pourtant, certains philosophes ont déclaré que franchir cette barrière était strictement impossible. Thomas Nagel (1974, 1986) a soutenu qu'il n'y avait pas moyen de passer du niveau objectif de la physiologie au niveau subjectif de la phénoménologie. Plus récemment, Colin Mc Ginn a soutenu que la conscience avait une « structure cachée » qui se tient à la fois derrière la phénoménologie et la physiologie, et que bien que cette structure cachée pourrait combler le fossé, elle nous est probablement à jamais inaccessible.

« Le type de structure cachée que j'envisage ne figurerait à aucun des niveaux envisagés par Nagel. Elle serait située quelque part entre eux. Ni phénoménologique, ni physique, ce niveau intermédiaire ne serait pas (par définition) conçu sur le modèle de l'un ou l'autre des côtés du fossé, et par conséquent ne se retrouverait pas lui-même incapable d'atteindre l'autre côté. Le caractériser appellerait une innovation conceptuelle radicale, dont j'ai soutenu qu'elle était probablement hors de notre portée » (Mc Ginn 1991, p. 102-103).

Le niveau de description du « logiciel » ou de la « machine virtuelle » que j'ai exploité dans ce livre est exactement la sorte de niveau intermédiaire que Mc Ginn décrit : il n'est pas explicitement physiologique ou mécanique, et pourtant il est capable de fournir les ponts nécessaires vers la machinerie du cerveau d'un côté, tout en étant de l'autre pas explicitement phénoménologique, tout en étant néanmoins capable de fournir les ponts nécessaires en direction du monde du contenu, les mondes de l'hétérophénoménologie. Nous y sommes arrivés ! Nous *avons* bien imaginé comment un cerveau pourrait produire de l'expérience consciente. Pourquoi Mc Ginn pense-t-il qu'il est hors de notre portée de s'engager dans cette « innovation conceptuelle radicale » ? Soumet-il les diverses approches de l'esprit en termes de logiciel à une analyse rigoureuse et détaillée qui démontre leur futilité ? Non. Il ne les examine absolument pas. Il n'essaie

même pas d'imaginer les niveaux intermédiaires qu'il postule : il se contente de noter que cela lui semble évident qu'il n'y a rien à espérer de ce côté-là.

La soi-disant « évidence » est un grand obstacle au progrès dans la compréhension de la conscience. C'est la chose la plus naturelle du monde que de penser la conscience comme se produisant dans une sorte de Théâtre Cartésien, et de supposer qu'il n'y a rien de véritablement problématique dans cette façon de penser. Cela paraît évident jusqu'à ce que vous preniez le temps de vous intéresser à ce que nous pourrions apprendre au sujet des activités du cerveau, et que vous commenciez à essayer d'imaginer, en détail, un modèle rival. Ce qui se produit alors ressemble assez à l'effet que produit le fait d'apprendre comment un magicien de théâtre réalise l'un de ses trucs. Une fois que nous examinons sérieusement ce qui se passe dans les coulisses, nous découvrons que nous ne voyions pas en fait ce que nous pensions se trouver sur le devant de la scène. L'énorme fossé qui sépare la phénoménologie et la physiologie se réduit quelque peu ; nous voyons que certains des traits « évidents » de la phénoménologie ne sont en définitive pas réels : il n'y a pas de remplissement avec du figment ; il n'y a pas de qualia intrinsèques ; il n'y a pas de fontaine centrale de sens et d'action, il n'y a pas de Théâtre Cartésien ; la distinction même entre les processus du devant de la scène et ceux qui se passent en coulisses perd son attrait. Il nous reste encore quantité de phénomènes étonnants à expliquer, mais quelques-uns des effets les plus époustoufflants n'existent tout simplement pas, et par conséquent ne requièrent pas d'explication.

À partir du moment où nous avons fait quelques progrès dans cette tâche ardue, en imaginant *comment* un cerveau produit les phénomènes de la conscience, nous nous mettons à faire quelques ajustements dans la tâche facile : imaginer que quelqu'un ou quelque chose est conscient. Nous pouvons continuer à penser ce phénomène en postulant une sorte de flux de conscience, mais nous ne donnons plus à ce flux ses propriétés traditionnelles. À présent que le flux de la conscience a été repensé sous la forme des opérations d'une machine virtuelle réalisée dans le cerveau, il cesse d'être « évident » que nous sommes victimes d'une illusion quand nous imaginons qu'un tel flux pourrait se loger dans le cerveau électronique d'un robot.

Mc Ginn invite ses lecteurs à se rendre avec lui : il est tout

simplement impossible d'imaginer comment du logiciel pourrait rendre un robot conscient. Pas la peine d'essayer, nous dit-il. D'autres philosophes ont promu cette attitude en inventant des expériences de pensée qui « marchent » précisément parce qu'elles dissuadent le lecteur d'essayer d'imaginer, en détail, comment le logiciel pourrait y parvenir. Curieusement, les deux expériences de pensée de ce type les plus connues font allusion à la Chine : l'expérience de pensée de la Nation Chinoise de Ned Block (1978), et celle de la Chambre Chinoise de John Searle (1980, 1982, 1984, 1988)¹. Ces deux expériences de pensée reposent sur la même fausse direction qu'on impose à l'imagination, et puisque l'expérience de Searle a été la plus discutée, je me concentrerai sur celle-ci. Searle nous invite à imaginer qu'il est enfermé dans une pièce, manipulant un programme d'IA géant, supposé comprendre le chinois. Il stipule que le programme passe avec succès le test de Turing, trompant toutes les tentatives des interlocuteurs humains pour le distinguer d'un individu authentiquement capable de comprendre le chinois, ou d'une conscience chinoise, dans la Chambre Chinoise. Searle, enfermé dans la pièce, est occupé à manipuler les suites de symboles du programme d'après les instructions du programme, mais il ne gagne pas de ce fait la moindre capacité de compréhension du chinois, et il n'y a rien d'autre non plus dans la pièce qui puisse comprendre le chinois (c'est « tout simplement évident », comme le dirait Frank Jackson).

Cette expérience de pensée est supposée prouver l'impossibilité de ce que Searle appelle l'« IA forte », la thèse selon laquelle « un ordinateur digital correctement programmé avec les bonnes entrées et les bonnes sorties aurait par là même un esprit exactement au sens où les êtres humains ont des esprits » (Searle, 1988a). Les diverses versions de cette expérience de pensée qu'a fournies Searle ont fait couler une quantité d'encre considérable pendant les dix dernières années, et bien que les philosophes et d'autres aient toujours trouvé des défauts dans l'expérience de pensée quand on la considère comme un argument logique², il est indéniable que sa « conclusion » continue à paraître « évidente » aux yeux de nombreuses personnes. Pourquoi ? Parce que les gens n'imaginent pas la situation assez dans le détail, comme il le faudrait.

Voici une expérience informelle qui nous aidera à voir pourquoi mon diagnostic est correct. En premier lieu, imaginons un

bref extrait tiré du dialogue qui a emporté la palme avec le juge du test de Turing pour l'expérience de la Chambre Chinoise (par commodité, j'ai traduit le chinois en français).

— JUGE : Avez-vous entendu parler de l'Irlandais qui avait trouvé une lampe magique ? Quand il l'a frottée un génie est apparu et lui a accordé trois vœux. « Je veux une pinte de Guinness ! » répondit l'Irlandais, et immédiatement elle apparut. L'Irlandais se mit à la siffler et à l'avaler goulûment, mais le verre de Guinness était toujours magiquement plein. Au bout d'un moment le génie s'impatienta. « Bon. Quel était ton second vœu ? » demanda-t-il. Entre deux lampées, l'Irlandais répondit : « Oh ! ben, je pense que j'aimerais bien avoir une autre pinte comme ça ! »

— CHAMBRE CHINOISE : Très drôle. Non, en fait, je la connaissais déjà — mais vous savez, je trouve que les blagues ethniques sont de mauvais goût. J'ai ri malgré moi, mais en fait je trouve que vous devriez trouver d'autres sujets de discussion.

— J : Soit, mais je vous ai raconté la blague parce que je voulais que vous me l'expliquiez.

— CC : La barbe ! On ne doit jamais expliquer les histoires drôles.

— J : Néanmoins, c'est ma question test. Pouvez-vous m'expliquer comment et pourquoi cette histoire « marche » ?

— CC : Si vous insistez. Vous voyez, cela dépend de l'hypothèse que le verre qui se remplit par magie se remplira à nouveau éternellement, en sorte que l'Irlandais a toute la bière qu'il puisse jamais vouloir boire. Il peut donc difficilement avoir besoin d'un autre verre magique mais il est à ce point stupide (c'est la partie de l'histoire qui me gêne) et imbibé d'alcool qu'il ne s'en aperçoit pas ; c'est pourquoi, extasié que son premier vœu se soit réalisé, il réclame une autre pinte de bière. Ces hypothèses d'arrière-plan ne sont pas vraies, bien entendu, mais font seulement partie de l'atmosphère ambiante des histoires drôles, dans laquelle nous suspendons notre croyance en l'impossibilité de la magie, et ainsi de suite. Entre parenthèses, nous pourrions imaginer une suite quelque peu laborieuse dans laquelle l'Irlandais se trouverait finalement avoir eu « raison » en formulant son second vœu — peut-être entend-il organiser une grande fête, et un seul verre ne se reremplira pas assez vite pour satisfaire tous ses hôtes assoiffés (et cela ne sert à rien de la conserver d'avance — nous savons tous comment la bière éventée perd sa saveur). Nous tendons à ne pas penser à ce genre de complications, qui fait partie de la façon dont les histoires drôles fonctionnent. Cela vous suffit-il ?

Cette conversation n'est pas éblouissante, mais supposons qu'elle suffise à tromper le juge. À présent, nous sommes invités à

imaginer que tous ces discours de CC sont composés par le programme géant que Searle stimule diligemment à la main. Difficile à imaginer ? Bien sûr, mais puisque Searle stipule que le programme passe le test de Turing, et puisque ce niveau de complexité conversationnelle serait certainement à sa portée, à moins que nous essayions d'imaginer les complexités d'un programme qui serait capable d'engendrer ce genre de conversation, nous ne suivons pas ses indications. Bien entendu, nous pourrions aussi imaginer que Searle n'a pas la moindre idée de ce qu'il est en train de faire dans la Chambre Chinoise ; il ne voit que des séries de zéros et de uns qu'il manipule conformément au programme. Au passage, il est important que Searle nous invite à imaginer qu'il manipule des caractères chinois incompréhensibles plutôt que des zéros et des uns, parce que cela peut nous induire, sans que nous prenions garde, à supposer (sans preuves) que le programme géant pourrait fonctionner simplement en « alignant » les caractères chinois d'entrée en face de certains caractères chinois de sortie. Aucun programme de ce genre ne pourrait fonctionner, bien entendu – est-ce que les paroles de CC en français « s'alignent » sur les questions du juge ?

Un programme qui pourrait effectivement engendrer les paroles de CC en réponse aux questions de J pourrait, si on le mettait en marche, avoir la forme suivante (du point de vue du niveau de la machine virtuelle, pas du point de vue primaire adopté par Searle). En analysant les premiers mots (« Avez-vous entendu parler... »), certains des démons détecteurs d'histoires drôles du programme ont été activés, et ils ont mis en marche une quantité de stratégies appropriées au traitement de la fiction, du langage « de seconde intention », et ainsi de suite, de sorte que quand les mots « lampe magique » ont été analysés, le programme avait déjà une priorité faible sur des réponses protestant qu'il n'y a pas de lampes magiques. Divers cadres narratifs d'histoires de génies dans des lampes (Minsky, 1975) ou de scripts (Schank et Abelson, 1977) ont été activés, créant diverses attentes de suites, mais celles-ci ont été court-circuitées, en fait, par la chute de l'histoire, qui invoquait un script plus ordinaire (le script pour « en redemander », et le caractère inattendu de cette chute ne fut pas perdu sur le programme... En même temps, des démons sensibles aux connotations négatives des récits d'histoires drôles ethniques furent également alertés, ce qui conduisit finalement

au second thème de la première réponse de CC... et ainsi de suite, avec bien plus de détails que je n'ai essayé d'en esquisser ici.

Le fait est que tout programme qui pourrait effectivement tenir sa place dans la conversation qu'on vient de décrire devrait être un système extraordinairement puissant, sophistiqué et à de nombreux niveaux, regorgeant de « connaissance du monde », de métaconnaissances et de métamétaconnaissances sur ses propres réponses, sur les réponses probables de son interlocuteur, sur ses propres « motivations » et sur celles de son interlocuteur, et bien d'autres choses encore. Searle ne nie pas que les programmes puissent avoir cette structure, cela va de soi. Il nous décourage simplement de nous en occuper. Mais si nous sommes supposés bien imaginer le cas qu'on nous présente, nous ne sommes pas simplement autorisés à, mais obligés d'imaginer que le programme que Searle stimule a toute cette structure – et plus encore, si seulement nous pouvions l'imaginer. Mais si c'est le cas, on ne comprend plus l'histoire drôle. *Peut-être* que les milliards d'actions de toutes ces parties hautement structurées produisent finalement une authentique compréhension dans le système. Si votre réponse à cette hypothèse est que vous n'en avez pas la moindre idée, c'est déjà suffisant pour montrer que l'expérience de pensée de Searle dépend, illicitement, du fait que vous imaginiez un cas trop simple, un cas non pertinent, et que vous en tiriez la conclusion « évidente ».

Voici comment on se trouve emmené dans la mauvaise direction. Nous voyons clairement que s'il y avait de la compréhension dans un tel système géant, ce ne serait pas la compréhension de Searle (puisque'il n'est qu'un rouage dans la machinerie, ignorant du contexte de ce qu'il est en train de faire). Nous pouvons aussi voir clairement qu'il n'y a rien qui ressemble de loin à de la compréhension dans un bout de programme trop petit pour être imaginé facilement – quoi qu'il puisse être, ce n'est qu'une routine stupide pour transformer des suites de symboles en d'autres suites de symboles selon une recette mécanique ou syntaxique. C'est alors que vient la prémisse qu'on a supprimée : sans doute *une plus grande quantité de la même chose*, peu importe combien, ne pourra jamais revenir à de la compréhension authentique. Mais pourquoi devrait-on croire que c'est vrai ? Les dualistes cartésiens le croiraient, parce qu'ils pensent que même les cerveaux humains sont incapables de parvenir à une compréhension par eux-mêmes ; selon la thèse cartésienne, il faut une âme

immatérielle pour produire le miracle de la compréhension. Si, d'un autre côté, nous sommes des matérialistes convaincus que d'une façon ou d'une autre nos cerveaux sont seuls responsables, sans assistance miraculeuse, de notre compréhension, nous devons admettre qu'une compréhension authentique est en quelque façon le produit d'un processus composé d'interactions entre une quantité de sous-systèmes dont aucun ne comprend une chose par lui-même. L'argument qui commence par : « cette petite quantité d'activité cérébrale ne comprend pas le chinois, pas plus que cette plus vaste quantité dont elle fait partie... » va droit à la conclusion indésirable selon laquelle même l'activité du cerveau tout entier est insuffisante pour rendre compte de la compréhension du chinois. Il est *difficile d'imaginer* comment « une plus grande quantité de la même chose » pourrait finir par revenir à de la compréhension, mais nous avons de bonnes raisons de croire que c'est possible, et par conséquent dans ce cas nous devrions renouveler nos efforts, et non pas abandonner.

Comment pourrions-nous renouveler nos efforts ? Avec l'aide de quelques concepts commodes : les concepts du logiciel de niveau intermédiaire que les informaticiens ont précisément forgés pour nous permettre de garder le contact avec des détails complexes de vastes systèmes, détails qui seraient, sans ces concepts, impossibles à imaginer. À ce niveau intermédiaire, nous pouvons discerner beaucoup d'entités qui sont tout à fait invisibles à un niveau plus microscopique, comme les « démons » rencontrés plus haut, à qui on attribue une forme de compréhension réduite. À partir de là, il ne devient pas si difficile d'imaginer comment « plus de la même chose » pourrait revenir à une authentique compréhension. Tous ces démons ainsi que d'autres entités sont organisés en un système énorme, dont les activités s'organisent elles-mêmes autour de son propre Centre de Gravité Narrative. Searle, alors qu'il travaille dans la Chambre Chinoise, ne comprend pas le chinois, mais il n'est pas seul dans la pièce. Il y a aussi le Système, CR, et c'est à ce moi-là que nous devrions attribuer la compréhension de l'histoire drôle.

Cette réponse à l'exemple de Searle est ce qu'il appelle la Réponse du Système. Cela a été la réponse usuelle des chercheurs en IA dès les premières versions de cette expérience de pensée, il y a plus de dix ans, mais les gens qui n'appartiennent pas à la communauté des chercheurs en IA l'apprécient rarement. Pourquoi ? Probablement parce qu'ils n'ont pas appris comment ima-

giner un tel système. Ils ne peuvent tout simplement pas imaginer comment la compréhension pourrait être une propriété qui émerge à partir de quantités de quasi-compréhensions distribuées au sein d'un vaste système. Ils n'y parviennent certainement pas s'ils n'essaient pas, mais comment pourrait-on les aider dans cet exercice difficile ? Est-ce qu'on « triche » quand on considère le logiciel comme composé d'homoncules qui quasi-comprennent, ou bien est-ce précisément le bon auxiliaire de notre imagination pour lui permettre de donner un sens à cette complexité astronomique ? Searle fait une pétition de principe. Il nous invite à imaginer que le programme géant consiste en une architecture simple de lecture de tableaux qui aligne directement des suites de caractères chinois avec d'autres, comme si ce genre de programme pouvait, en toute justice, être vraiment tenu pour un programme. Nous n'avons aucun mal à imaginer ce genre de programme simple ni à supposer que ce soit le programme que Searle manipule, puisque aucun programme de ce style ne pourrait produire des résultats susceptibles de passer le test de Turing, comme il a été annoncé. (Pour une réponse semblable, et sa critique, voir Block, 1982, et Dennett, 1985.)

La complexité est une chose importante. Si elle ne l'était pas, il y aurait un argument bien plus bref contre l'IA forte : « Hé, regardez cette calculatrice de poche. Elle ne comprend pas le chinois, et tout ordinateur qu'on puisse concevoir n'est qu'une calculatrice de poche géante, par conséquent aucun ordinateur n'est capable de comprendre le chinois. CQFD. » Quand nous prenons en compte la complexité, comme nous devons le faire, nous devons réellement la prendre en compte, et pas seulement prétendre le faire. C'est difficile, mais tant que nous ne le faisons pas, toute intuition que nous puissions avoir quant à ce qui n'est « évidemment » pas présent doit être suspecte. Comme le cas de Marie la spécialiste des couleurs, l'expérience de pensée de Searle ne produit une conviction forte, claire, que si nous ne suivons pas les instructions. Ces pompes à intuition sont défectueuses ; elles ne stimulent pas notre imagination, et au contraire l'envoient sur une voie de garage.

Mais alors, et mes propres pompes à intuitions ? Que dire de Shakey le robot, ou du CADAVEUGLE de Série I, ou du patient à vision aveugle entraîné par biofeedback, par exemple ? Ne sont-elles pas elles aussi suspectes, elles aussi coupables de fourvoyer le lecteur ? J'ai certainement fait de mon mieux pour conduire

vosre imagination sur certaines voies, et pour vous empêcher de vous empêtrer dans des complications que je ne jugeais pas nécessaires pour les besoins de ma démonstration. Il y a quelque asymétrie, cependant : mes pompes à intuition sont, pour la plupart, destinées à vous aider à imaginer des possibilités nouvelles, pas à vous convaincre de l'existence d'impossibilités de principe. Ma variation sur le thème du cerveau dans la cuve qui a ouvert le livre était destinée à vous faire sentir l'impossibilité de certaines sortes de tromperies et certaines expériences de pensée du chapitre 4 étaient destinées à montrer que, à moins qu'il n'y ait un Théâtre Cartésien, il *ne pourrait pas* y avoir de fait décisif qui permette de distinguer les révisions de contenu orwelliennes des révisions staliniennes. Mais ces expériences de pensée reposaient sur une tentative pour rendre les choses plus nettes à l'« opposition » ; ainsi l'exemple de la femme au chapeau dans la soirée et celui de la femme aux cheveux longs à lunettes étaient destinés à renforcer l'intuition que je cherchais à discréditer par mon argumentation.

N'empêche, que le lecteur prenne garde : mes pompes à intuition, comme celles de quiconque, ne sont pas les démonstrations en bonne et due forme qu'elles peuvent sembler être ; elles relèvent plus de l'art que de la science. (Pour d'autres mises en garde au sujet des expériences de pensée des philosophes, voir Wilkes, 1988.) Si elles nous aident à concevoir de nouvelles possibilités, que nous pouvons ensuite confirmer par des méthodes plus systématiques, c'est un succès ; si elles nous envoient sur une voie de garage, c'est navrant. Même de bons outils peuvent être mal utilisés, et comme n'importe quel autre ouvrier, nous aurions intérêt à savoir comment nos outils fonctionnent.

Quel effet cela fait d'être une chauve-souris ?

L'expérience de pensée portant sur la conscience la plus citée et la plus influente est celle de Thomas Nagel : « Quel effet cela fait d'être une chauve-souris ? » (1974 ³). Il répond à cette question en soutenant qu'il nous est impossible de l'imaginer. Beaucoup de gens, apparemment, trouvent que cette affirmation coule de source ; on voit quelquefois son article cité par des savants comme si c'était la rareté des raretés, un « résultat » philoso-

phique – une démonstration établie d'un fait que toute théorie ultérieure doit exploiter.

Nagel a bien choisi ses créatures de référence. Les chauves-souris, en tant que mammifères comme nous, sont suffisamment semblables à nous pour nous convaincre que *bien entendu* elles sont conscientes. (S'il avait écrit « Quel effet cela fait d'être une araignée ? » bien des gens auraient été enclins à se demander ce qui le rendait si sûr que cela fasse un quelconque effet.) Mais grâce à leur système d'écholocation – les chauves-souris peuvent « voir avec leurs oreilles » – elles sont assez différentes de nous pour que nous puissions percevoir le fossé énorme qui les sépare de nous. S'il avait écrit un article intitulé « Quel effet cela fait d'être un chimpanzé ? », ou, pour en venir au fait, « Quel effet cela fait d'être un chat ? » l'opinion selon laquelle sa conclusion pessimiste est évidente n'aurait pas fait autant l'unanimité. Il y a beaucoup de gens qui sont totalement sûrs de savoir *exactement* quel effet cela fait d'être un chat. (Ils ont tort, bien entendu, à moins qu'ils n'aient associé toutes leurs observations pleines d'amour et d'empathie à de grandes quantités de recherches physiologiques ; mais ils se fourvoieraient alors complètement, du point de vue de Nagel.)

Quoi qu'il en soit, la plupart des gens semblent tout à fait prêts à prouver le « résultat » de Nagel quant à l'inaccessibilité pour nous de la conscience des chauves-souris. Certains philosophes l'ont contesté, pourtant, et à juste titre (Hofstadter, 1981 ; Hardin, 1988 ; Leiber, 1988, Akins, 1990). Il nous faut tout d'abord nous mettre au clair sur la nature du résultat en question. Ce n'est pas l'affirmation épistémologique ou empirique selon laquelle même si quelqu'un réussissait (« par accident ») à imaginer quel effet cela fait d'être une chauve-souris, il ne serait jamais capable de confirmer que cet exploit d'imagination s'est bien produit. C'est plutôt que nous autres humains n'avons pas et n'acquerrons jamais les moyens, la machinerie représentationnelle, nécessaires pour nous représenter quel effet cela fait d'être une chauve-souris.

La distinction est importante. Au chapitre 11 nous avons envisagé un exploit similaire : imaginer quel effet cela fait d'être un habitant de Leipzig écoutant l'une des cantates de Bach pour la première fois. Le problème épistémologique est difficile, mais on peut le traiter avec les techniques usuelles. Se figurer quelles sortes d'expériences ils pourraient avoir eues, et comment celles-

ci auraient différé de nos propres expériences de la musique de Bach, est une affaire de recherche historique, culturelle, et peut-être physiologique. Nous pouvons nous figurer une partie de celles-ci très aisément, y compris certaines des différences les plus frappantes par rapport à notre propre expérience, mais si nous essayons de nous placer dans la série elle-même des états d'expériences qu'une telle personne aurait éprouvés, nous obtiendrions des rendements décroissants. La tâche exigerait que nous nous soumettions à des transformations énormes – oublier la majeure partie de ce que nous savons, perdre des associations et des habitudes, en acquérir de nouvelles. Nous pouvons utiliser notre recherche « à la troisième personne » pour dire ce que seraient ces transformations, mais les éprouver réellement impliquerait qu'on paie le prix terrible de l'isolement par rapport à notre propre culture – ne pas écouter la radio, ne pas lire d'informations sur les développements politiques et sociaux ultérieurs à l'époque de Bach, etc. On n'a pas besoin d'en passer par là pour apprendre des choses sur la conscience d'un habitant de Leipzig.

Il en est de même quand il s'agit d'imaginer quel effet cela fait d'être une chauve-souris. Nous devrions nous intéresser à ce que nous pouvons savoir de la conscience de la chauve-souris (si elle en a une), pas à la question de savoir si nous pouvons transformer momentanément nos esprits en des esprits de chauves-souris. Au chapitre 11, nous avons mis à mal le pré-supposé qu'il existe des propriétés « intrinsèques » – des qualia – qui constitueraient *quel effet cela fait* d'avoir une expérience consciente ou une autre, et comme le fait remarquer Akins (1990), même s'il y avait des propriétés résiduelles non dispositionnelles, non relationnelles des expériences ressenties par les chauves-souris, le fait de se trouver intimement en contact avec celles-ci, tout en demeurant ignorant des faits accessibles scientifiquement quant à la structure systématique de la perception et du comportement des chauves-souris. Il y a au moins une grande quantité de choses que nous pouvons savoir quant à l'effet que cela fait d'être une chauve-souris, et ni Nagel ni qui que ce soit d'autre ne nous ont donné une raison importante de croire qu'il y ait quoi que ce soit d'important ou de théoriquement important qui nous soit inaccessible.

Nagel soutient qu'aucune quantité de connaissance à la troisième personne ne pourrait nous dire quel effet cela fait d'être une chauve-souris, et je rejette purement et simplement cette

affirmation. Comment pourrions-nous vider cette querelle ? En nous engageant dans quelque chose qui a l'air au départ d'être un jeu enfantin – un jeu dans lequel une personne imagine quel effet cela fait d'être x , et où l'autre essaie de démontrer qu'il y a quelque chose de faux dans cet exercice particulier d'hétérophénoménologie.

Voici quelques exercices pour nous échauffer :

– A. : Voici Winnie l'ours en peluche, qui pense que ce serait bien agréable d'avoir du miel pour son petit déjeuner !

– B. : Faux. L'ours en peluche n'est pas en mesure de distinguer le miel de quoi que ce soit d'autre. Pas d'organes des sens, pas même d'estomac. L'ours en peluche est rempli de rembourrage inerte, cela ne fait aucun effet d'être un ours en peluche.

– A. : Voici Bambi le daim, admirant le joli coucher de soleil, jusqu'à ce que le beau ciel orange lui rappelle la veste du méchant chasseur !

– B. Faux. Les daims sont aveugles aux couleurs (bon, ils ont une espèce de vision chromatique). Même si les daims peuvent être conscients, ils ne distinguent pas des couleurs comme l'orange.

– A. Voici Jiji la chauve-souris percevant, avec son système spécial de sonar, que la chose volante qui plonge vers lui n'était pas sa cousine Kiki, mais un aigle, les plumes de ses ailes déployées et les griffes prêtes à tuer !

– B. Arrête – à quelle distance as-tu dit qu'était l'aigle ? L'écholocation de la chauve-souris ne marche qu'à quelques mètres.

– A. Hum, bon... Et l'aigle n'était qu'à deux mètres !

– B. Ah, maintenant, c'est plus difficile à dire. Quelles sont exactement les limites de résolution de l'écholocation de la chauve-souris ? Sert-elle à identifier les objets ou n'est-ce qu'un système d'alarme et de détection pour la capture ? Une chauve-souris serait-elle capable de distinguer des plumes déployées de plumes non déployées simplement en utilisant l'écholocation ? J'en doute, mais il nous faut encore monter quelques expériences pour le savoir, et aussi, bien entendu, quelques expériences pour découvrir si les chauves-souris sont capables de détecter, de réidentifier, les autres membres de leur espèce. Certains mammifères le peuvent, et d'autres, on l'a montré, sont entièrement ignorants de ces faits.

Les recherches que suggère cet exercice nous entraîneraient loin dans une analyse de la structure du monde perceptif et

comportemental de la chauve-souris. Nous pourrions ainsi classer les récits hétérophénoménologiques par degré de plausibilité, en rejetant ceux qui asserteraient ou présupposeraient des capacités discriminatives, ou des dispositions réactives que l'écologie et la neurophysiologie de la chauve-souris ne justifieraient manifestement pas. Par exemple, nous apprendrions que les chauves-souris ne s'inquiètent pas des cris perçants qu'elles émettent pour produire leurs échos, parce qu'elles ont un muscle très bien conçu qui ferme leurs oreilles en synchronisation parfaite avec ces cris, un peu comme les dispositifs de synchronisation qui permettent aux systèmes de radars sensibles d'éviter d'être anéantis par leurs propres signaux. Il y a beaucoup de travaux sur ces sujets, et nous en savons déjà bien plus, par exemple sur la raison pour laquelle les chauves-souris utilisent diverses structures de fréquences pour leurs cris, selon qu'elles détectent une proie, approchent une cible, ou s'apprêtent à tuer (Akins, 1989, 1990).

Quand nous parvenons à produire des récits hétérophénoménologiques incontestables, nous devrions les accepter – provisoirement, en attendant d'autres découvertes – en tant qu'explications correctes de l'effet que cela fait d'être la créature en question. C'est, après tout, ainsi que nous nous traitons entre nous, entre humains. Quand je recommande que nous traitions les chauves-souris et d'autres êtres à interpréter de la même façon, je ne suis pas en train de *déplacer* le poids de la preuve sur d'autres entités.

Nous pourrions utiliser ces recherches pour démystifier toutes sortes d'illusions totalement romantiques quant à la conscience des chauves-souris. Nous *savons* que le délicieux livre pour enfants de Randall Jarrett, *The Bat Poet* (1963), est une fiction, parce que nous savons que les chauves-souris ne parlent pas ! Des affirmations moins nettement fantastiques portant sur la phénoménologie laissent place à des faits moins évidents, mais néanmoins publics, quant à leur physiologie et à leur comportement. Les recherches nous montreraient beaucoup de choses sur ce dont une chauve-souris pourrait être consciente ou ce dont elle ne pourrait pas être consciente dans diverses circonstances, en nous montrant quelles ressources elles ont dans leurs systèmes nerveux pour représenter ceci ou cela, et en vérifiant expérimentalement si la chauve-souris a bien mis l'information en question au service de la modulation de son comportement. Il est difficile d'imaginer ce que l'on peut tirer de ce genre de recherches tant qu'on ne

s'y met pas. (Pour une analyse détaillée surprenante de l'effet que cela fait d'être un singe vervet, voir Cheney et Seyfarth, *How Monkeys See the World*, 1990.)

Tout cela invite à faire l'objection évidente suivante. Ces recherches nous montreraient beaucoup de choses sur l'organisation du cerveau et du traitement de l'information de la chauve-souris, mais elles nous montreraient seulement ce dont les chauves-souris ne sont *pas* conscientes, et elles laisseraient totalement ouverte la question de savoir de quoi les chauves-souris *sont* conscientes (si elles le sont). Comme nous le savons, une bonne partie du traitement de l'information dans les systèmes nerveux est inconsciente, si bien que ces méthodes d'investigation ne feraient rien qui permette d'exclure l'hypothèse que les chauves-souris pourraient être... des zombies volants, des créatures que cela ne fait aucun effet d'être ! (Wilkes, 1988, p. 224, se demande si l'écholocalisation est une sorte de vision aveugle, qui ne fait aucun effet subjectif.)

Ah, la chauve-souris est sortie du sac. Cette discussion semble s'engager dans cette direction de mauvais augure, et nous devons continuer tout droit. Richard Dawkins (1986), dans une analyse éclairante de l'organisation naturelle des chauves-souris *Rhinolophes fer à cheval*, donne une version très claire de l'image qui commence à se faire jour :

« L'effet Doppler est utilisé dans les contrôles radar qui permettent à la police de piéger les automobilistes... En comparant la fréquence de l'onde émise avec celle de l'écho de la police, *ou plutôt son enregistreur automatique* (mes italiques), on peut calculer la vitesse de chaque voiture... En comparant la hauteur de son cri avec la hauteur de l'écho, donc, la chauve-souris (*ou plutôt l'ordinateur embarqué dans son cerveau* (mes italiques)) pourrait en théorie calculer à quelle vitesse elle se rapproche de l'arbre » (p. 30-31, tr. fr., p. 46-47).

Il est tentant de demander : y a-t-il quelque chose dans la chauve-souris qui soit situé relativement à son « ordinateur embarqué » (qui opère sans le moindre soupçon de conscience) comme la police est située relativement à leur « enregistreur automatique » ? Les policiers n'ont pas à calculer le déplacement Doppler consciemment, mais il leur faut faire l'expérience, consciemment, du cadran sur leur enregistreur qui dit, en symboles LED lumineux : « 150 à l'heure ». C'est ce qui les fait sauter sur leurs motocyclettes et mettre en marche leurs sirènes. Nous

pouvons supposer à bon droit que la chauve-souris elle aussi ne calcule pas consciemment le déplacement Doppler – son ordinateur embarqué s'en charge – mais alors y a-t-il encore un rôle, dans la chauve-souris, pour quelque chose comme le flic qui éprouve une expérience, un témoin pour apprécier (consciemment) la « sortie » de l'analyse-de-l'effet-Doppler par ordinateur ? Notez bien que nous pourrions assez facilement remplacer l'officier de police par un dispositif automatique qui enregistrerait d'une façon quelconque le numéro de la plaque du véhicule en infraction, trouvait le nom et l'adresse du conducteur et lui enverrait la contravention. Il n'y a rien de spécial dans la tâche de la police qui montre que cela ne pourrait pas se faire sans la moindre expérience de quoi que ce soit. Il en est de même, semble-t-il, pour la chauve-souris. Une chauve-souris pourrait être un zombie. Elle serait un zombie – c'est ce que ce raisonnement suggère – sauf s'il y avait un observateur interne en elle qui réagirait à une présentation innée un peu comme les officiers de police réagissent aux lumières qui clignotent sur leurs instruments.

Ne tombez pas dans le piège. C'est notre vieille némésis, l'Auditoire dans le Théâtre Cartésien. *Votre* conscience ne consiste pas dans le fait que votre cerveau soit habité par un agent interne à qui votre cerveau présente des informations, et c'est pourquoi si nous n'étions pas capables de trouver un tel agent central dans le cerveau de la chauve-souris cela ne menacerait pas la possibilité qu'elle ait une conscience, ou la possibilité pour nous dire à quoi peut ressembler son expérience consciente. Pour comprendre la conscience d'une chauve-souris, nous devons simplement appliquer les mêmes principes à la chauve-souris que ceux que nous nous appliquons à nous-mêmes.

Mais que pourrait faire, par conséquent, une chauve-souris qui soit suffisamment spéciale pour nous convaincre que nous sommes en présence d'une conscience authentique ? Il peut sembler que même si nous inventons des utilisateurs-de-sorties situés derrière le transducteur-Doppler de la chauve-souris, et quels qu'ils soient, il n'y aurait aucune raison convaincante, externe, en « troisième personne », d'admettre que la chauve-souris a une expérience consciente. Mais ce n'est pas le cas. Si la chauve-souris pouvait parler, par exemple, elle produirait un texte à partir duquel nous pourrions produire un monde hétérophénoménologique, et cela nous donnerait exactement la même raison de

lui accorder la conscience que nous pourrions attribuer à n'importe quelle personne. Mais comme nous l'avons noté à l'instant, les chauves-souris ne peuvent pas parler. Elles peuvent, cependant, avoir de nombreux comportements non verbaux qui permettent clairement d'ébaucher une description de leur monde hétérophénoménologique, ou comme J. von Exhül, le pionnier de ces recherches, l'a dit, leur *Umwelt und Innenwelt*, leur Monde Environnant et leur Monde Intérieur.

Il n'est pas impossible de pratiquer l'hétérophénoménologie sans texte ; c'est seulement difficile (Dennett, 1988a, 1989a, 1989b). Il existe une branche de l'hétérophénoménologie appelée éthologie cognitive, qui consiste à modéliser l'esprit animal en étudiant – expérimentalement – leur comportement sur le terrain. Les possibilités et les difficultés de ce genre d'enquête sont bien représentées dans Cheney et Seyfarth (1990), Whiten and Byrne (1988) et dans Ristau (1991), un recueil d'hommages en l'honneur de Donald Griffin, le pionnier de l'analyse de l'écho-localisation chez les chauves-souris et le créateur de l'éthologie cognitive. L'une des difficultés frustrantes à laquelle ces chercheurs se trouvent confrontés est que beaucoup des expériences que l'on rêve de monter se révèlent totalement impossibles à réaliser en l'absence de langage ; on ne parvient tout simplement pas à *mettre en place* les sujets (et personne ne les a mis en place) de la façon qui serait requise par ces expériences sans converser avec eux (Dennett, 1988a).

Ce n'est pas simplement un problème épistémologique qui se pose à l'hétérophénoménologue : la difficulté même qu'on a à créer les circonstances expérimentales requises dans l'environnement naturel démontre quelque chose de plus fondamental quant aux esprits des créatures sans langage. Cela montre que les situations écologiques de ces animaux ne leur ont jamais fourni les *possibilités* de développer (par évolution, apprentissage, ou les deux) un grand nombre des activités mentales avancées qui façonnent nos esprits, et c'est pourquoi nous pouvons être tout à fait sûrs qu'ils ne les ont jamais développées. Par exemple, pensez au concept de *secret*. Un secret n'est pas seulement quelque chose que vous savez et que les autres ne savent pas. Pour que vous ayez un secret il vous faut savoir que les autres ne le savent pas, et il vous faut être capable de contrôler ce fait. (Si vous êtes le premier à voir la débandade de la troupe, vous pouvez savoir quelque chose que les autres ne savent pas, mais pas pour long-

temps ; vous ne pouvez pas *garder* secrète cette information privilégiée.) L'écologie comportementale d'une espèce doit être structurée de façon spéciale pour que les secrets puissent jouer un rôle. Les antilopes, dans leurs bandes, n'ont pas de secrets et aucun moyen d'en obtenir. C'est pourquoi une antilope n'est probablement pas plus capable d'ourdir un plan secret qu'elle n'est capable de compter jusqu'à cent ou d'admirer les couleurs d'un coucher de soleil. Les chauves-souris, qui s'engagent dans des raids relativement solitaires pendant lesquels elles pourraient être capables de reconnaître cet isolement par rapport à leurs rivales, ne satisfont aucune des conditions nécessaires pour avoir des secrets. Ont-elles également des intérêts détectables qui pourraient être servis en exploitant des secrets ? (Que pourrait faire une palourde avec un secret ? Juste s'asseoir là dans la boue, en riant sous cape ?) Les chauves-souris ont-elles aussi des habitudes de ruse ou de tromperie dans leurs activités de chasse qui pourraient être adaptées pour des activités plus élaborées de conservation de secrets ? Il y a en fait de nombreuses questions de ce genre qui, une fois posées, suggèrent d'autres enquêtes et d'autres expériences. La structure de l'esprit d'une chauve-souris est tout aussi accessible que la structure de son système digestif : pour analyser l'un ou l'autre, il faut faire des va-et-vient systématiques entre une analyse de ses contenus et une analyse du monde dont ses contenus sont dérivés, en faisant attention aux méthodes et aux objectifs de la dérivation.

Wittgenstein a dit un jour « Si un lion pouvait parler, nous ne pourrions pas le comprendre » (1958, p. 223). Je pense, au contraire, que si un lion pouvait parler, ce lion aurait un esprit si différent de l'esprit usuel des lions que, bien que le comprenant fort bien, nous n'apprendrions pas grand-chose de lui sur les lions ordinaires. Le langage, comme nous l'avons vu dans les chapitres antérieurs, joue un rôle énorme dans la structuration d'un esprit humain. On ne devrait pas supposer que l'esprit d'une créature sans langage – qui n'a pas vraiment besoin de langage – soit structuré de la même façon. Cela veut-il dire que les animaux sans langage « ne sont pas conscients du tout » (comme le soutenait avec force Descartes) ? Cette question soulève toujours à ce moment une sorte de défi incrédule, mais nous ne devrions pas nous sentir obligés d'y répondre sous cette forme. Notez qu'elle présuppose quelque chose que nous avons eu du mal à éviter : l'hypothèse que la conscience est une sorte de propriété

spéciale, qui est affaire de tout ou rien, et divise l'univers en deux catégories extrêmement différentes : les choses qui l'ont (les choses pour lesquelles cela fait un certain effet d'être ces choses, pour le dire comme Nagel) et les choses qui ne l'ont pas. Même dans notre propre cas, nous ne pouvons pas tracer la ligne de partage entre nos propres états conscients et nos états inconscients. La théorie de la conscience que nous avons esquissée permet de nombreuses variations dans l'architecture fonctionnelle, et bien que la présence du langage marque un accroissement particulièrement dramatique en capacités imaginatives, en versatilité, en contrôle (pour ne mentionner que quelques-uns des pouvoirs évidents d'une machine virtuelle joycienne), ces pouvoirs n'ont pas besoin d'avoir *en plus* le pouvoir d'allumer une lumière intérieure quelconque qui serait sinon éteinte.

Quand nous imaginons quel effet cela fait d'être une créature sans langage, nous partons, naturellement, de notre propre expérience, et la majeure partie de ce qui nous vient à l'esprit doit être ajusté (principalement à la baisse). La sorte de conscience qu'éprouvent les animaux est dramatiquement tronquée, comparée à la nôtre. Une chauve-souris, par exemple, non seulement ne peut pas se demander si nous sommes aujourd'hui vendredi, mais elle ne peut même pas se demander si elle est une chauve-souris ; il n'y a pas de rôle pour l'activité de jeu dans sa structure cognitive. Alors qu'une chauve-souris, comme n'importe quel humble homard, a un moi biologique, elle n'a pas d'ego égoïque dont elle puisse parler – pas de Centre de Gravité Narrative, ou si elle en a un, il est négligeable. Pas de mots au-bout-de-salange, mais aussi pas de regrets, pas d'aspirations complexes, pas de réminiscences nostalgiques, pas de grands projets, pas de réflexions sur quel effet cela fait d'être un chat, ou même sur quel effet cela fait d'être une chauve-souris. Cette liste de rejets serait du scepticisme à bon compte si nous n'avions pas une théorie empirique positive pour le fonder. Est-ce que je prétends avoir prouvé que les chauves-souris ne *pourraient* pas avoir ces états mentaux ? Eh bien, non, mais je ne peux pas non plus prouver que les champignons *ne pourraient pas* être des vaisseaux intergalactiques qui nous espionnent.

N'est-ce pas un préjugé terriblement anthropocentrique ? Qui plus est, que dire des sourds et muets ? Ne sont-ils pas conscients ? Bien sûr qu'ils le sont – mais n'en tirons pas des conclusions extravagantes sur leur conscience, par sympathie mal placée.

Quand un sourd et muet acquiert le langage (en particulier le langage Signe, le langage le plus naturel qu'un sourd et muet puisse apprendre), un esprit humain complet est né, nettement et manifestement différent de l'esprit d'une personne dont l'ouïe est normale, mais capable de toute la complexité réflexive et du pouvoir génératif – peut-être de plus encore. Mais sans un langage naturel, l'esprit d'un sourd et muet est terriblement bloqué (voir Sacks, 1989, en particulier la bibliographie annotée). Comme le philosophe Ian Hacking (1990) le remarque dans un compte rendu du livre de Sacks, « il faut beaucoup d'imagination même pour avoir le sentiment de ce que n'a pas l'enfant sourd ». On ne fait pas aux sourds et muets une faveur en imaginant qu'en l'absence de langage, ils sont capables d'éprouver tous les délices mentaux que nous autres humains entendants pouvons éprouver, et on ne fait pas une faveur aux animaux non humains en essayant d'obscurcir les faits que nous connaissons quant aux limitations de leurs esprits.

Et cela, comme nombre d'entre vous êtes impatients de le faire remarquer, est un sous-texte qui s'est battu pendant un moment pour parvenir à la surface : beaucoup de gens ont peur de voir la conscience expliquée parce qu'ils ont peur que si nous réussissions à l'expliquer, nous perdrons nos statuts moraux. Peut-être *pouvons-nous* imaginer un ordinateur conscient (ou la conscience d'une chauve-souris), mais *nous ne devrions pas essayer*, pensent-ils. Si nous prenons cette mauvaise habitude, nous commencerons à traiter les animaux comme s'ils étaient des jouets à ressorts, les bébés et les sourds et muets comme s'ils étaient des ours en peluche et – pour ajouter l'insulte à la blessure – les robots comme s'ils étaient vraiment des gens.

Matière à penser

J'emprunte le titre de cette section (*Minding and Mattering*) à un article de Marian Stamp Dawkins (1987), qui a fait des enquêtes approfondies sur les implications morales de l'hétérophénoménologie animale (ses travaux antérieurs figurent dans son livre *Animal Suffering : The Science of Animal Welfare*, 1980). Comme elle le note, nos attitudes morales envers les autres animaux sont pleines d'incohérences.

« Il nous suffit de penser aux différentes sortes d'animaux pour faire surgir nos incohérences. Il y a des manifestations contre le meurtre des bébés phoques, mais il n'y a pas de campagnes comparables pour arrêter le meurtre des rats. Beaucoup de gens sont très heureux de manger des cochons ou des moutons, mais sont horrifiés à l'idée de manger des chiens ou des chevaux » (p. 150).

Dawkins fait remarquer qu'il y a deux points de vue dans cet embrouillamini : l'aptitude à raisonner et l'aptitude à souffrir. Descartes exploitait beaucoup l'inaptitude des animaux non humains à raisonner (tout au moins de la façon dont les animaux humains raisonnent), ce qui lui a valu une réponse fameuse de la part du philosophe utilitariste Jeremy Bentham : « Un cheval ou un chien adultes sont sans comparaison des animaux plus rationnels et plus aptes à ce qu'on leur adresse la parole qu'un nouveau-né d'un jour ou deux, ou même d'un mois. Mais supposez qu'ils soient autres qu'ils ne sont, qu'est-ce que cela nous prouverait ? La question n'est pas : peuvent-ils *raisonner* ? ni : peuvent-ils *parler* ? mais : peuvent-ils *souffrir* ? » (Bentham, 1789). Ces deux points de vue semblent être les deux critères opposés du statut moral, mais comme le soutient Dawkins, « donner une valeur éthique à la capacité de souffrir nous conduira en dernière analyse à valoriser les animaux qui sont habiles. Même si nous rejetons d'entrée de jeu le critère cartésien du raisonnement, il s'agit du raisonnement des animaux qui sont les plus probablement susceptibles de posséder la capacité de souffrir » (p. 153).

L'explication est implicite dans la théorie de la conscience que nous avons développée. Souffrir, ce n'est pas être visité par un état ineffable mais intrinsèquement horrible, c'est avoir ses espoirs de vie, ses plans de vie, ses projets d'existence ruinés par des circonstances imposées sur vos désirs, bloquant vos intentions – quelles qu'elles soient. L'idée que la souffrance puisse s'expliquer en quelque façon par la présence d'une propriété intrinsèque quelconque – l'horribilité, disons – est aussi vaine que la tentative d'explication du rire par la présence d'une hilarité intrinsèque. Par conséquent, l'inaccessibilité présumée, l'inconnaissabilité ultime, de la souffrance de quelqu'un d'autre, est tout aussi trompeuse que les autres fantaisies portant sur les qualia intrinsèques que nous avons démasquées, bien que son caractère pernicieux soit plus évident. Il s'ensuit – et cela s'accorde avec notre intuition – que la capacité à souffrir est fonction de la capacité

à avoir des désirs, des attentes, articulés, à long terme, très précis, et d'autres états mentaux sophistiqués.

Les êtres humains ne sont pas les seuls êtres qui soient assez intelligents pour souffrir ; le cheval et le chien de Bentham montrent par leur comportement qu'ils ont assez de complexité mentale pour distinguer – et pour s'inquiéter de – une gamme de douleurs et d'autres épreuves qui est loin d'être négligeable, même si c'est une gamme étroite, comparée aux possibilités de la souffrance humaine. D'autres animaux, en particulier les singes, les éléphants, et les dauphins, ont des registres apparemment bien plus vastes.

Pour compenser le fait qu'ils ont à endurer toutes les souffrances, les êtres intelligents doivent avoir tout l'amusement. Il faut avoir une économie cognitive avec un budget suffisant d'exploration et d'automanipulation pour fournir l'espace nécessaire aux magasins récurifs de désirs dérivés qui rendent l'amusement possible. Un être a franchi une première étape quand son architecture lui permet d'apprécier la signification de « Arrête ! j'adore ça ! » Des versions superficielles de ce pouvoir de construction sont manifestes dans quelques espèces supérieures, mais cela prend beaucoup d'imagination, et de loisir – quelque chose que la plupart des espèces ne peuvent pas se permettre – pour faire intervenir une gamme importante de plaisirs. Plus la portée est vaste, plus des détails sont riches, plus les désirs sont précis ; le pire c'est quand ces désirs sont contrariés.

Mais vous avez une question sur les lèvres : pourquoi serait-il important que les désirs d'une créature soient contrariés si ce ne sont pas des désirs *conscients* ? Réponse : pourquoi cela serait-il plus important s'ils étaient conscients – en particulier si la conscience était une propriété, comme le pensent certains, qui échappe à jamais à toute analyse ? Pourquoi les espoirs détruits d'un « zombie » comptent-ils moins que les espoirs détruits d'une personne consciente ? Il y a ici un tour de passe-passe qu'il faut exposer et rejeter. La conscience, dites-vous, est ce qui compte, mais dans ce cas, vous vous en tenez à des doctrines au sujet de la conscience qui nous empêchent systématiquement de tirer avantage des *raisons* pour lesquelles elle compte. Postuler l'existence de qualités internes qui ne soient pas seulement privées et qui aient une valeur intrinsèque, mais également impossibles à mettre en évidence et à attester, est tout simplement de l'obscurantisme.

Dawkins montre comment on peut analyser expérimentalement les différences attestables – les seules qui comptent vraiment – et il vaut la peine de donner quelques détails pour voir combien on peut tirer de choses d'expériences simples avec une espèce plutôt défavorisée :

« Les poules gardées au-dehors ou dans de vastes litières passent une bonne partie de leur temps à gratter le sol alentour, ce qui m'a fait suspecter que l'absence de litière dans les cages des batteries pourrait les faire souffrir. Sans doute, quand je leur ai donné à choisir entre une cage avec un sol grillagé et une autre avec de la litière, elles choisirent la cage avec la litière. En fait, elles entraient dans une petite cage (si petite qu'elles pouvaient difficilement y tourner) si c'était la seule façon pour elles d'avoir accès à la litière. Même des oiseaux qui ont été emprisonnés toute leur vie dans des cages et qui n'avaient jamais eu l'expérience antérieure de la litière choisirent la cage avec de la litière sur le sol. C'était encourageant, mais cela ne suffisait pas. Il me fallait montrer que non seulement les poules ont une préférence pour la litière mais aussi qu'elles avaient une préférence qui était suffisamment forte pour qu'elles souffrent si on les en privait.

J'offris alors aux poules un choix légèrement différent. Cette fois-ci elles eurent à choisir entre une cage à sol grillagé dans laquelle se trouvaient de la nourriture et de l'eau et une cage avec de la litière mais sans nourriture ni eau... Le résultat fut qu'elles passèrent un bon moment dans la cage avec la litière, et beaucoup moins de temps dans la cage grillagée, même si c'était le seul endroit où elles pouvaient se nourrir et boire. Puis j'introduisis une complication. Les oiseaux avaient à "travailler" pour se mouvoir entre les cages. Ils avaient soit à sauter d'un corridor soit à pousser un rideau de plastique noir. Passer d'une cage à l'autre avait à présent un coût... Les poules passèrent encore la même quantité de temps dans la cage grillagée avec la nourriture comme auparavant quand il n'y avait pas de difficulté à entrer dedans. Mais elles ne passèrent pratiquement pas de temps dans la cage à litière. Elles semblaient simplement ne pas être prêtes à travailler ou à payer un prix quelconque pour entrer dans la cage à litière... Contrairement à ce à quoi je m'attendais, les oiseaux semblaient dire que la litière ne comptait pas vraiment pour eux » (p. 157-159).

Elle conclut qu'« on peut mettre en évidence une souffrance de l'esprit émotionnel chez les animaux qui ont assez d'esprit rationnel pour être capables d'intervenir sur les conditions qui les font souffrir », et elle ajoute qu'« il est aussi probable que les organismes qui n'ont pas la capacité de faire quoi que ce soit

pour s'écarter d'une source de danger ne développeraient pas évolutivement une capacité à souffrir. Du point de vue évolutif, cela ne servirait à rien qu'un arbre dont on coupe les branches ait la capacité de souffrir en silence » (p. 159). Comme nous l'avons vu au chapitre 6 (voir aussi chapitre 2, note 6), il faut être prudent quand on forge ce genre d'arguments évolutionnistes en termes de fonction, car l'histoire joue un grand rôle dans l'évolution, et l'histoire peut jouer des tours. Mais en l'absence de justifications positives pour imputer de la souffrance, ou de justifications positives de suspecter que ces justifications positives sont pour une raison quelconque systématiquement cachées, nous devrions conclure qu'il n'y a pas de souffrance. Nous n'avons pas besoin de craindre que cette règle austère nous conduise à relâcher nos obligations envers nos frères animaux. Elle nous fournit toujours des justifications largement suffisantes pour tirer des conclusions positives : beaucoup, sinon la plupart, des animaux sont capables de souffrir à un degré significatif. On défendra bien mieux l'idée qu'il faut bien traiter les animaux en admettant les vastes différences de degré, qu'en promouvant pieusement un insupportable dogme quant à l'universalité et l'égalité de la douleur animale.

Cela peut régler la question objective qui porte sur la présence ou l'absence de souffrance, mais cela ne calme pas les sentiments moraux inquiets qu'on éprouve à l'idée d'expliquer la conscience d'une façon aussi mécanique et insensible. L'enjeu est plus important.

J'ai une ferme dans le Maine, et j'adore l'idée qu'il y ait des ours et des coyotes qui vivent dans mes bois. Je les vois rarement et je ne vois même pas de signes de leur présence, mais j'aime simplement l'idée qu'ils sont là, et je serais malheureux d'apprendre qu'ils sont partis. Je ne sentirais pas de compensation de cette perte si l'un de mes amis chercheur en IA remplissait mes bois de quantités de bestioles robots (bien que l'idée, si on l'imagine dans le détail, soit plaisante). C'est important pour moi qu'il y ait des créatures sauvages, descendant de créatures sauvages, qui vivent si près de moi. De même cela me fait plaisir qu'il y ait des concerts qu'on joue dans la région de Boston que non seulement je n'écoute pas, mais dont je n'entendrai jamais parler.

Ce sont là des faits d'un type particulier. Ce sont des faits qui nous importent simplement parce qu'une des parties de notre

environnement qui compte pour nous est celui de nos croyances. Et puisque nous ne nous laissons pas facilement bernier à continuer à croire des propositions une fois que les raisons que nous avons de les croire ont disparu, cela compte pour nous que les croyances soient *vraies*, même si nous ne voyons pas directement pour quelles raisons elles peuvent l'être. Comme toute autre partie de l'environnement, un environnement de croyances peut être fragile, composé de parties interconnectées à la fois par des accidents historiques et des liens bien organisés. Pensez, par exemple, à cette partie délicate de notre environnement de croyances qui porte sur la disposition de nos corps après la mort. Peu de gens croient que l'âme réside dans le corps après la mort – même les gens qui croient aux âmes ne croient pas une telle chose. Et pourtant, bien peu d'entre nous, peut-être même personne, toléreraient une « réforme » qui encouragerait les gens à se débarrasser de leurs parents morts en les mettant dans des sacs à ordures en plastique, ou en s'en débarrassant d'une autre façon peu cérémonieuse. Pourquoi ? Pas parce que nous croyons que les cadavres souffrent effectivement d'une quelconque indignité. Un cadavre ne peut pas plus souffrir d'une indignité qu'une bûche. Et pourtant l'idée est choquante, repoussante. Pourquoi ?

Les raisons sont complexes, mais on peut en donner quelques-unes ici. Une personne n'est pas seulement un corps ; une personne *a* un corps. Ce cadavre est le corps de notre cher vieux Dupond, un Centre de Gravité Narrative qui doit sa réalité tout autant à nos efforts coopératifs d'interprétation hétérophénoménologique mutuelle qu'au corps qui est à présent sans vie. Les limites de Dupond ne sont pas identiques aux limites du corps de Dupond, et les intérêts de Dupond, grâce à la pratique humaine curieuse de tissage du soi, peuvent s'étendre au-delà des intérêts élémentaires qui ont promu cette pratique. Nous traitons son corps avec respect parce que c'est important pour la préservation de l'environnement de croyances dans lequel nous vivons tous. Si nous commençons à traiter les cadavres comme des ordures, par exemple, cela pourra changer la façon dont nous traiterons de quasi-cadavres – ceux qui sont toujours en vie mais mourants. Si nous ne péchons pas par souci de prolonger les rituels et les pratiques de respect bien au-delà du seuil de la mort, les mourants (et ceux qui s'occupent d'eux) subiront une angoisse, un affront, une *possibilité* qui risque de les offenser. Traiter « mal » un cadavre peut ne pas faire de mal directement à une personne

mourante, et ne fait certainement aucun mal au cadavre lui-même ; mais, si cela devenait pratique courante, et si cela se savait (ce qui ne manquerait pas d'être le cas), cela changerait de façon significative l'environnement de croyances qui entoure les morts. Les gens imagineraient les événements qui seraient supposés suivre leur départ différemment de la façon qu'ils les imaginent aujourd'hui, et de façons particulièrement déprimantes. Peut-être pas pour une bonne raison, mais qu'importe. Si les gens risquent d'en être déprimés, c'est une bonne raison pour ne pas adopter une politique.

Il existe donc des raisons indirectes, mais pourtant fiables, légitimes, fortes, pour continuer à respecter les cadavres. Nous n'avons pas besoin d'une mythologie supposant quelque chose de spécial qui résiderait en fait dans les cadavres et qui leur donnerait un privilège. Ce *pourrait* être un mythe utile qu'on pourrait répandre parmi les gens peu sophistiqués, mais ce serait montrer une extrême condescendance que de penser que nous sommes parmi les mieux informés et qu'à ce titre il nous incombe de préserver ces mythes. De même, il existe d'excellentes raisons de traiter tous les animaux sauvages avec soin et sollicitude. Ces raisons sont relativement indépendantes des faits qui portent sur le degré de souffrance que peuvent avoir les animaux de telle ou telle espèce. Ils dépendent plus directement du fait que diverses croyances sont ambiantes dans notre culture, et comptent pour nous, qu'il soit vrai ou non qu'elles *devraient* compter pour nous. Puisqu'elles comptent présentement, elles comptent. Mais la rationalité de l'environnement de croyances — le fait que des croyances stupides ou infondées tendent à s'éteindre à la longue — en dépit de la superstition — implique que les choses qui comptent à présent peuvent ne pas compter toujours.

Mais alors, comme nous l'avons entrevu au chapitre 1, une théorie qui attaque de front l'environnement général de croyances a la possibilité de causer de la souffrance (chez les gens qui s'inquiètent particulièrement des animaux par exemple, quoi qu'il arrive aux animaux revient à ce qu'ils souffrent). Cela veut-il dire que nous devrions supprimer toute analyse de ces problèmes, de peur d'ouvrir la boîte de Pandore ? Cela pourrait se justifier, si nous pouvions nous convaincre du fait que notre environnement de croyances présentes, mangé aux mythes ou pas, était de toute évidence un environnement moralement acceptable, bénin. Mais je soutiens qu'il est clair que ce n'est pas le

cas. Ceux qui s'inquiètent de ce que peuvent coûter ces révélations non demandées devraient jeter un œil sur les coûts des mythes courants. Pensons-nous vraiment que ce à quoi nous sommes confrontés en ce moment vaille la peine que nous le protégeions d'un quelconque obscurantisme créatif ? Ne pensons-nous pas, par exemple, que l'on devrait laisser de côté de vastes ressources pour préserver la possibilité imaginaire de renouveler la vie mentale des gens qui sont dans le coma, alors que nous n'avons pas de ressources à dépenser pour améliorer le sort des espoirs désespérés, mais loin d'être imaginaires, des pauvres ? Les mythes qui portent sur le caractère sacré de la vie, ou de la conscience, sont à double tranchant. Ils peuvent être utiles pour ériger des barrières (contre l'euthanasie, contre la peine capitale, contre l'avortement, contre la consommation de viande) pour impressionner les gens qui manquent d'imagination, mais chez ceux qui disposent de quelques lumières au prix d'une hypocrisie offensante ou au prix d'un aveuglement volontaire ridicule.

Les barrières absolutistes, comme la ligne Maginot, jouent rarement le rôle qu'elles sont supposées jouer. La campagne qu'on a menée à une époque contre le matérialisme a déjà succombé à l'embarras de ses promoteurs, et la campagne contre l'« IA forte », bien qu'en elle-même aussi pleine de bonnes intentions, ne peut offrir que la théorie de l'esprit la plus élimée. Il vaudrait sûrement mieux essayer de promouvoir un intérêt pour des justifications *non* absolutistes, *non* intrinsèques et *non* fondées sur des dichotomies de celles de nos préoccupations morales qui peuvent coexister avec l'augmentation de nos connaissances des mécanismes internes de cette machine étonnante, le cerveau. Les arguments moraux *des deux côtés* sur les questions de la peine capitale, de l'avortement, de la consommation de viande, et de la vivisection, par exemple, s'élèveront d'un cran si nous essayons de rejeter les mythes qui n'ont de toute façon pas besoin qu'on les protège.

La conscience expliquée ou éliminée ?

Quand nous apprenons que la seule différence qu'il y a entre l'or et l'argent est le nombre de particules sub-atomiques contenues dans leurs atomes, nous pouvons nous sentir trompés ou irrités – les physiciens ont éliminé quelque chose : le doré de

l'or ; ils ont exclu l'argenté de l'argent, que nous apprécions. Et quand ils expliquent la façon dont la réflexion et l'absorption de la radiation électromagnétique rendent compte des couleurs et de la vision des couleurs, ils semblent négliger la chose même qui compte le plus. Mais bien entendu il doit y avoir « quelque chose qu'on laisse de côté » – sinon nous n'aurions pas commencé à expliquer. Laisser quelque chose de côté n'est pas simplement une caractéristique des explications ratées, c'en est une aussi des explications réussies.

Seule une théorie qui expliquerait les événements conscients en termes d'événements inconscients pourrait expliquer la conscience. Si notre modèle de la façon dont la douleur est un produit de l'activité cérébrale contient toujours une boîte appelée « douleur », vous n'avez pas encore commencé à expliquer ce qu'est la douleur, et si votre modèle de la conscience se poursuit jusqu'au moment magique où il vous faut dire « alors se produit un miracle », vous n'avez pas commencé à donner l'ombre d'une explication de la nature de la conscience.

Cela conduit certaines personnes à soutenir que la conscience ne pourra jamais être expliquée. Mais pourquoi la conscience serait-elle la seule chose que l'on ne pourrait expliquer ? Les solides, les liquides et les gaz peuvent être expliqués en termes de choses qui ne sont pas elles-mêmes des solides, des liquides ou des gaz. Il ne fait pas de doute que la vie puisse s'expliquer en termes de choses qui ne sont pas elles-mêmes vivantes – et l'explication ne rend pas les choses vivantes sans vie. L'illusion que la conscience est une exception vient, je le soupçonne, d'une impossibilité à comprendre ce trait général de toute explication réussie. En pensant, à tort, que l'explication laisse quelque chose de côté, nous pensons sauver ce qui serait autrement perdu en le remettant dans l'observateur comme quale – ou sous la forme de quelque autre propriété « intrinsèquement » mystérieuse. Le psychisme devient la jupe protectrice sous laquelle tous ces jolis petits chatons adorables peuvent se cacher. Il peut y avoir des *motifs* pour penser que la conscience ne peut pas être expliquée, mais j'espère que je l'ai montré, il y a de bonnes *raisons* de penser qu'elle peut l'être.

Mon explication de la conscience est loin d'être complète. On pourrait même dire que ce n'était qu'un début, mais *c'est un début*, parce que cela rompt le charme du cercle enchanté d'idées qui ont rendu l'explication de la conscience impossible en appa-

rence. Je n'ai pas remplacé une théorie métaphorique, le Théâtre Cartésien, par une théorie *non* métaphorique (« littérale », « scientifique »). Tout ce que j'ai fait, c'est remplacer une famille de métaphores et d'images par une autre, en échangeant le théâtre, le Témoin, le Signifieur Central, le Figment, contre le Logiciel, les Machines Virtuelles, les Versions Multiples, un Pandémonium d'Homuncules. Ce n'est qu'une guerre de métaphores ; les métaphores sont les outils de la pensée. Personne ne peut réfléchir sur la conscience sans elles. C'est pourquoi il vous faut vous doter des meilleurs outils disponibles. Pourriez-vous l'avoir imaginée sans eux ?

NOTES

1. Essayez d'imaginer l'état d'esprit de Ji Hu-Min, mon étudiant de doctorat originaire de Pékin, qui, dans un cours d'introduction à la philosophie de l'esprit anglo-américaine (alors que son anglais était encore rudimentaire), participait à un séminaire où les étudiants et les professeurs débattaient vigoureusement sur le point de savoir ce qui se passerait si la population entière de la Chine était forcée d'une certaine façon à participer à la réalisation massive d'un programme d'IA présumé conscient (c'est l'exemple de Block), pour ensuite discuter, avec toujours aussi peu de respect des sensibilités d'un observateur chinois, le cas de la Chambre Chinoise de Searle.

2. La réfutation définitive, à laquelle Searle n'a toujours pas répondu de façon adéquate, est celle de Douglas Hofstadter dans Hofstadter et Dennett (1981), p. 273-382. Au cours des années il y a eu bien d'autres critiques tranchantes. Dans « Pensée rapide » (dans Dennett, 1987a) j'ai proposé un nouveau diagnostic des sources de confusion dans cette expérience de pensée. Sa réponse a consisté à déclarer, sans argument à l'appui, que tous les points avancés dans cet article étaient dénués de pertinence (Searle, 1988b). Aucun prestidigitateur n'aime voir ses trucs expliqués au public.

3. *What it is like to be a bat*. Voir la remarque du traducteur dans la version française.

Appendice A (pour philosophes)

Il y a certains endroits dans le livre où je glisse rapidement sur des batailles philosophiques majeures, ou bien où, sous une autre forme, je manque de façon éclatante à tous mes devoirs de philosophe académique. Les philosophes qui ont lu mon manuscrit ont soulevé des questions sur ces failles, questions méritant des réponses, mais qui portent sur des sujets qui peuvent ne pas intéresser les non-philosophes.

Question – *Vous semblez aller à toute vitesse à la fin du chapitre 10, dans le dialogue avec Otto, lorsque vous introduisez brièvement les « pressentiments » comme s'il s'agissait d'actes de langage dans lesquels il n'y aurait ni Acteur ni Discours ; vous revenez alors sur votre propre caricature, en remplaçant les pressentiments par des « événements de fixation de contenu » sans autre forme d'explication. Est-ce qu'il ne s'agit pas là du tournant crucial de toute votre théorie ?*

Réponse – Effectivement. C'est le point de contact premier avec l'autre moitié de ma théorie de l'esprit, la théorie du contenu ou de l'intentionnalité très récemment présentées dans *La Stratégie de l'interprète*. Il y a maints autres endroits dans le livre où je m'appuie sur cette théorie, mais vous avez localisé le point qui, je crois, a le plus de poids. Sans cette théorie du contenu, ma théorie devrait dire à cet endroit : « Et alors un miracle se produit. » Ma stratégie fondamentale a toujours été la même : d'abord développer une analyse du contenu qui soit *indépendante de* et *plus fondamentale que* la conscience – une analyse du contenu qui traite pareillement de toute fixation de contenu inconscient (dans les cerveaux, dans les ordinateurs, dans la

« recognition » par l'évolution de propriétés d'organisations sélectionnées) – et en second lieu, élaborer sur ces bases une analyse de la conscience. D'abord, le contenu, ensuite la conscience. Les deux moitiés de *Brainstorms* récapitulaient la stratégie, mais au fur et à mesure que les moitiés de la théorie se développaient, elles ont dépassé la taille d'un seul volume. Ce livre complète ma troisième exécution de la campagne. Cette stratégie est complètement opposée, naturellement, à la vision de Nagel et de Searle, qui, chacun à sa manière, veulent absolument considérer que la conscience est fondationnelle. La raison pour laquelle j'ai glissé si rapidement sur cette question absolument centrale au chapitre 10 est simplement que je ne voyais pas comment on pouvait de manière utile télescoper les centaines de pages d'analyse et d'argumentation que j'avais consacrées à la théorie du contenu de manière à en faire quelque chose qui soit à la fois précis et accessible. Donc, si vous trouvez que je suis allé à toute vitesse dans ces pages, je vous prie de consulter la version lente dans les autres pages citées en bibliographie.

Q. – *Il semble, malgré tout, qu'il y ait une tension – sinon une plate contradiction – entre les deux moitiés de votre théorie. La posture intentionnelle présuppose (ou nourrit) la rationalité, et donc l'unité de l'agent – le système intentionnel –, alors que le modèle des Versions Multiples s'oppose tout au long à cette Unité Centrale. Quelle est, selon vous, la bonne manière de concevoir un esprit ?*

R. – Cela dépend complètement de la distance à laquelle vous êtes parvenu. Plus vous êtes près, plus la désunion, la multiplicité et la compétitivité paraissent importantes. La source principale du mythe du Théâtre Cartésien, après tout, c'est l'extrapolation paresseuse de la posture intentionnelle, *tout du long*. Traiter une entité complexe et mouvante comme un agent doté d'un seul esprit est une manière magnifique de voir la structure dans toute l'activité ; cette tactique nous vient naturellement, et se trouve même probablement génétiquement favorisée comme manière de percevoir et de penser. Mais quand nous aspirons à une science de l'esprit, il nous faut apprendre à restreindre et à réorienter ces habitudes de pensée, en décomposant l'agent à l'esprit unique en mini-agents et macro-agents (et sans aucun Chef unique). Il nous est alors possible de voir que bien des phénomènes *apparents* de l'expérience consciente sont mal décrits par la tactique unitaire traditionnelle. Les amortisseurs de choc qui ont affaire à la

tension sont les identifications forcées de questions relevant de l'hétérophénoménologie (conçues d'après la perspective traditionnelle) avec des événements qui relèvent de la fixation de contenu dans le cerveau (conçus selon la nouvelle perspective).

Les philosophes ont souvent attiré l'attention sur les idéalizations de la tactique traditionnelle, mais ils les ont moins souvent directement attaquées. C'est ainsi qu'on trouve une abondante littérature philosophique, consacrée aux difficultés de la logique des états réflexifs de croyance et de connaissance, qui commence avec Hintikka (1962). L'une des idéalizations essentielles de la formalisation de Hintikka, comme il l'a dit explicitement, c'était que les énoncés obéissant à la logique qu'il présentait « devaient se faire en *une seule et même occasion* [...] La notion d'oubli n'est pas applicable dans les limites d'une occasion » (p. 7). L'importance de cette limitation, notait-il, n'a pas toujours été appréciée – et il est caractéristique qu'elle se soit perdue dans les brumes de la controverse qui s'est ensuivie. Hintikka a reconnu que cette quantisation d'« occasions » est une simplification nécessaire requise pour formaliser les concepts quotidiens de croyance et de connaissance comme il l'a fait ; cela fixe le contenu à un instant donné, et fixe ainsi l'identité de la proposition en question. J'ai soutenu ici que cette individuation artificielle en « états » et en « temps » est l'un des traits qui transforment ces concepts hérités de la psychologie populaire en fantaisies imaginaires lorsque nous essayons de les faire cadrer avec les complexités de ce qui se passe dans l'esprit.

Q. – *Que sont finalement, selon vous, les expériences conscientes ? Êtes-vous un théoricien de l'identité, un matérialiste éliminativiste, un fonctionnaliste, un instrumentaliste ?*

R. – Je refuse de céder à la pression qui veut que je formule une proposition unique, formelle et proprement quantifiée qui exprimerait la ligne de force de ma théorie. Remplir la formule (*x*) (*x est une expérience consciente si et seulement si...*) et la défendre contre tel ou tel contre-exemple qu'on pourrait proposer n'est pas une bonne méthode pour développer une théorie de la conscience, et je pense avoir montré pourquoi. Le caractère indirect de la méthode hétérophénoménologique est précisément une manière de fuir les obligations immotivées d'« identification » ou de « réduction » des entités (hypothétiques) qui meublent l'ontologie des sujets. Est-ce que les anthropologues *identifient* Phénomène avec le type dont ils découvrent qu'il a accompli toutes

ces bonnes actions dans la jungle, ou sont-ils « éliminativistes » à l'égard de Phénhomme ? S'ils ont *bien* fait leur travail, la seule issue qui leur reste ne peut se décider que sur la base de la diplomatie, et non de la doctrine scientifique ou philosophique. À certains égards, vous pourriez dire que ma théorie identifie les expériences conscientes avec les événements qui portent l'information dans le cerveau – puisque c'est tout ce qui se passe, et que bien des événements du cerveau ont une ressemblance frappante avec les habitants des univers hétérophénoménologiques des sujets. Mais on pourrait juger « essentielles » d'autres propriétés des éléments hétérophénoménologiques – tels que la position que ces éléments prennent dans la suite temporelle subjective, auquel cas ils *ne pourraient pas* être identifiés avec les événements cérébraux qui se présentent, lesquels peuvent se trouver dans une autre suite, sous peine de violer la loi de Leibniz.

La question de savoir si l'on doit traiter une partie de l'univers hétérophénoménologique du sujet comme une fiction utile plutôt que comme une vérité un peu forcée n'est pas toujours une question qui mérite qu'on s'y attarde vraiment. Les images mentales sont-elles réelles ? Il y a des structures de données réelles dans le cerveau des gens qui ressemblent plutôt à des images – s'agit-il *là* des images mentales auxquelles vous faites allusion ? Si oui, alors la réponse est oui ; si non, alors la réponse est non. Les qualia sont-ils fonctionnellement définissables ? Non, parce qu'il n'existe pas de propriétés telles que les qualia. Ou bien, non, parce que les qualia sont des propriétés dispositionnelles des cerveaux qui ne sont pas strictement définissables en termes *fonctionnels*. Ou bien, oui, parce que si vous aviez vraiment tout compris du fonctionnement du système nerveux, vous comprendriez tout des propriétés auxquelles les gens font allusion lorsqu'ils prétendent parler des qualia.

Cela veut-il dire que je suis fonctionnaliste ? Oui et non. Je ne suis pas un fonctionnaliste des machines de Turing, mais alors je doute qu'il y en ait jamais eu un, ce qui est dommage, étant donné que tant de réfutations doivent alors être jetées au panier. Je suis évidemment une sorte de « téléofonctionnaliste », peut-être le premier téléofonctionnaliste (dans *Content and Consciousness*), mais comme je n'ai cessé de le dire de façon très claire, et l'ai souligné dans la discussion sur l'évolution et sur les qualia, je ne commets pas l'erreur qui consiste à essayer de définir toutes les différences mentales saillantes en termes de

fonctions biologiques. Ce serait faire une très mauvaise lecture de Darwin.

Suis-je instrumentaliste ? Je pense avoir montré dans « *Real Patterns* » (1991a) pourquoi c'est là une question mal conçue. Les douleurs sont-elles réelles ? Aussi réelles que des coupes de cheveux, des dollars, des occasions, des personnes, ou des centres de gravité, mais à quel point tout cela est-il réel ? Ces questions à structure dichotomique proviennent toutes de l'exigence que l'on se donne de remplir le blanc dans la formule quantifiée ci-dessus et certains philosophes pensent que l'on développe une théorie de l'esprit en concoctant une proposition de ce genre capable de parer à toutes les balles puis en la défendant. Une proposition unique n'est pas une théorie, c'est un slogan ; et ce que font certains philosophes, ce n'est pas de la théorie, c'est de l'affûtage de slogan. *Pour quoi* se donner toute cette peine ? Combien dissiperait-on de confusions, quel progrès ferait-on dans la manière de voir les choses si cette entreprise réussissait ? Avez-vous vraiment besoin d'imprimer quelque chose sur votre T-shirt ? Certains affûteurs de slogans sont très, très bons à cela, mais comme l'a un jour dit de façon mémorable le psychologue Donald Hebb, « si cela ne vaut pas le coup de le faire, cela ne vaut pas le coup de bien le faire ».

Non que je considère que la définition scrupuleuse et la critique des définitions au moyen de contre-exemples ne soit jamais un exercice valable. Prenons, par exemple, la définition de la couleur. Les récentes analyses et tentatives qui en ont été faites par les philosophes nous ont permis d'ouvrir les yeux. Elles ont vraiment mis en lumière les concepts et prévenu d'authentiques contresens. Étant donné le soin avec lequel les philosophes ont entrepris d'essayer de donner une définition précise de la couleur, eh bien, j'accorde que j'ai été rapide, exagérément rapide lorsque j'ai déclaré, au chapitre 12, que les couleurs sont des « propriétés réfléchissantes de la surface des objets, ou des volumes transparents ». *Quelles* propriétés réfléchissantes au juste ? Je crois avoir expliqué pourquoi essayer de répondre à cette question de manière précise serait perdre son temps ; la seule réponse précise ne serait pas une réponse concise, pour des raisons aisément compréhensibles. En d'autres termes, il est difficile de parvenir à une définition « non circulaire ». Mais encore ? Est-ce que je crois vraiment qu'en faisant ce simple pas, je parviendrai à faire face à tous les problèmes soulevés par mes concurrents, puis

résister aux questions soulevées par la compétition (auxquels j'ajouterai, en plus de ceux qui ont été cités auparavant, les noms de Strawson, 1989, et de Boghossian et Velleman, 1989, 1991) ? Oui, mais c'est une longue histoire, aussi laisserai-je la balle dans leur camp.

Q. — *Est-ce que votre position ne se ramène pas finalement à une sorte de vérificationnisme ?*

R. — Les philosophes ont récemment réussi à se convaincre — à convaincre maints spectateurs innocents — que le vérificationnisme est *toujours* un péché. Sous l'influence de Searle et de Putnam, par exemple, le neuroscientifique Gerald Edelman se dépêche de battre en retraite de crainte d'accomplir un acte de quasi-vérificationnisme. « L'absence de preuve de conscience de soi chez les animaux autres que les chimpanzés ne nous permet pas de considérer qu'ils ne sont pas conscients de soi » (1989, p. 280). Allons ! Courage ! Bien sûr que nous pouvons non seulement *considérer* qu'ils ne le sont pas, mais nous pouvons poursuivre l'examen, et si nous trouvons des raisons positives et fortes de le récuser, il nous faut le récuser. Il est temps d'inverser le mouvement du pendule. Dans un commentaire sur mes critiques antérieures de Nagel (Dennett, 1982a), Richard Rorty a dit un jour : « Dennett pense que l'on peut être sceptique sur l'intérêt insistant que porte Nagel aux vies intérieures phénoménologiquement riches des chauves-souris "sans devenir par là même le Vérificationniste du village". Je ne le pense pas. Je pense que le scepticisme que l'on peut avoir à l'égard d'intuitions comme celles de Nagel ou de Searle n'est plausible que s'il se fonde sur des considérations méthodologiques générales sur le statut des intuitions. Le reproche général du vérificationniste au réaliste est d'insister sur des différences (entre, par exemple, les chauves-souris qui ont une vie privée et celles qui n'en ont pas, entre les chiens qui ont une intentionnalité intrinsèque et ceux qui n'en ont pas), ce qui ne fait aucune différence : que l'on ne puisse intégrer ses intuitions dans un schème explicatif parce que ce sont des « rouages qui ne jouent aucun rôle dans le mécanisme » (Wittgenstein, 1953, I, § 271). Cela me paraît un bon reproche à faire, et le seul que nous ayons besoin de faire » (Rorty, 1982a, p. 342-343. Voir aussi Rorty, 1982b).

J'étais d'accord, mais j'ai proposé une légère (0,742) modulation de la thèse : « Sous les hourras du professeur Rorty... je suis prêt à sortir du placard sous la forme d'une sorte de véri-

ficationniste, mais, s'il vous plaît, pas un Vérificationniste du village ; soyons tous des vérificationnistes *urbains* » (Dennett, 1982b, p. 355). Ce livre va plus loin dans ce sens, et soutient que si nous ne sommes pas des vérificationnistes urbains, nous finirons par tolérer toutes sortes d'absurdités : l'épiphiénoménisme, les zombies, des spectres inversés indistinguables, des nounours conscients, des araignées conscientes de soi. L'endroit où se manifeste de la façon la plus frappante le genre de vérificationnisme que j'épouse se trouve au chapitre 4, lorsque j'essaie de développer une argumentation pour montrer que, puisqu'il n'y a *et ne pourrait y avoir* aucune preuve en faveur de modèles orwelliens ou staliniens de la conscience, il n'y a aucun fait décisif en la matière. Ce qu'on oppose en général à cette déclaration vérificationniste c'est que je préjuge du cours de la science ; comment puis-je savoir que de nouvelles découvertes dans les neurosciences ne *mettront pas au jour* de nouvelles raisons de faire cette distinction ? La réplique – qu'on n'entend pas beaucoup de nos jours – est vite trouvée : en ce qui concerne certains concepts (pas tous, mais certains) nous pouvons être certains d'en savoir assez pour savoir que quoi qu'on puisse découvrir un jour en matière de nouveauté scientifique, cela ne pourrait dégager ce genre de possibilité. Prenons, par exemple, l'hypothèse qui veut que l'univers soit à l'endroit, et sa négation, l'hypothèse que l'univers soit à l'envers. Est-ce que ces hypothèses sont en bonne posture ? Y a-t-il, ou pourrait-il y avoir ici un fait décisif en la matière ? Est-ce commettre un péché vérificationniste que d'être d'avis que, quelles que soient les révolutions cosmologiques qu'on nous propose, elles ne feront pas de *cette* « controverse » un fait empirique décisif qui permette de fixer les choses ?

Q. – *Mais vous êtes bien une sorte de behavioriste, non ?*

R. – On m'a déjà posé cette question, et je suis heureux de prendre à mon compte la réponse que Wittgenstein (1953) lui a donnée : « N'êtes-vous pas en réalité un behavioriste déguisé ? En vérité, n'êtes-vous pas au fond en train de dire que tout en dehors du comportement humain relève de la fiction ? » – Si je parle de fiction, alors il s'agit d'une fiction *grammaticale* (§ 307). Comment apparaît le problème philosophique concernant les processus et les états mentaux et le behaviorisme ? – La première étape est celle qui échappe complètement au regard. Nous parlons de processus et d'états et ne prenons pas parti sur leur nature. Peut-être un jour, pensons-nous, en saurons-nous un peu plus

sur eux. Mais c'est précisément ce qui nous amène à voir les choses d'une certaine façon. Car nous avons un concept bien déterminé de ce que veut dire apprendre à mieux connaître un processus. (On a accompli le pas décisif dans le tour de passe-passe, et c'était précisément celui qui nous semblait parfaitement innocent.) – Et voilà que l'analogie qui était censée nous faire comprendre nos pensées tombe en morceaux. Il nous faut donc nier le processus qui reste toujours incompris dans le médium qui reste toujours inexploré. Et il nous semble à présent que nous avons nié les processus mentaux. Ce que nous ne voulons évidemment pas faire (§ 308).

Plusieurs philosophes ont vu dans ce que je fais une sorte de reprise de l'attaque wittgensteimienne contre les « objets » de l'expérience consciente. C'est bien en effet de cela qu'il s'agit. Comme le paragraphe 308 le montre clairement, si nous voulons éviter le tour de passe-passe, il nous faut nous faire *d'abord* une idée de la « nature » des états et processus mentaux. C'est pourquoi j'ai mis neuf longs chapitres pour parvenir à un point où je puisse commencer à affronter les problèmes sous leur allure philosophique typique – bref sous leur fausse allure. Ma dette à l'égard de Wittgenstein est immense et ne date pas d'hier. Lorsque j'étais étudiant de première année, c'était mon héros, c'est pourquoi je suis allé à Oxford, où il paraissait être le héros de tout le monde. Quand je m'aperçus à quel point la plupart de mes condisciples étaient (à mon avis) complètement à côté de la plaque, je renonçai à essayer d'« être » un wittgensteinien, me contentant de prendre ce que je croyais avoir appris des *Investigations* pour essayer d'en tirer parti dans ma propre recherche.

Appendice B (pour scientifiques)

Les philosophes sont souvent accusés de se complaire dans la psychologie (ou les neurosciences ou la physique ou...) en fauteuil et les histoires embarrassantes foisonnent sur ces philosophes dont les confiantes déclarations *a priori* se sont vues par la suite réfutées dans les laboratoires. Une réponse raisonnable à ce risque établi est pour le philosophe de se replier prudemment sur ces arènes conceptuelles où il y a peu sinon aucun danger de jamais dire quelque chose qui puisse être infirmé (ou confirmé) par les découvertes empiriques. Une autre réponse raisonnable consiste à étudier, de son fauteuil, les meilleurs fruits du laboratoire, les meilleurs efforts des théoriciens qui s'appuient sur les bases solides de l'empirie, puis de poursuivre, avec sa philosophie pour seule arme, en essayant de mettre en lumière les obstacles, quitte à marcher occasionnellement à cloche-pied, dans l'intérêt d'y voir clair, d'une manière ou d'une autre, sur les implications de telle ou telle idée théorique. Quand on en vient à des questions conceptuelles, les savants ne sont pas plus à l'abri de la confusion que le profane. Après tout, les savants passent pas mal de temps dans leur fauteuil, essayant de se faire une idée sur la manière d'interpréter les résultats des expériences de tout un chacun, et ce qu'ils peuvent faire à ces moments-là se fond imperceptiblement dans ce que font les philosophes. Voilà qui est risqué, mais revigorant.

Voici donc quelques idées à moitié mûres pour des expériences conçues en vue de tester les implications du modèle de la conscience que j'ai esquissé. Je les ai choisies à partir d'un échantillon bien plus important d'idées au quart mûres qui soit n'ont pas

réussi à passer sous les fourches caudines de mes patients informateurs, soit ont déjà été mises en échec par eux. (Ma moyenne de points gagnants sur le dernier groupe est assez élevée pour m'encourager à poursuivre.) En tant que philosophe, j'ai essayé de conserver à mon modèle une allure aussi générale et peu compromettante que possible. Si j'ai fait ce qu'il fallait faire, ces expériences devraient aider à simplement établir jusqu'où au juste une version forte de mon modèle est confirmée ; si le modèle était entièrement infirmé, je serais bel et bien réfuté et embarrassé.

Sur le temps et la mesure du temps

Si une suite subjective est le produit d'une interprétation, et n'est pas directement fonction d'une véritable suite, il devrait être possible de créer de forts effets interprétatifs de toutes sortes qui soient indépendants de la véritable mesure du temps.

1. *La marche de l'araignée.* La lumière entre en contact selon une suite, mimant le lapin cutané, mais en cherchant à produire des jugements illusoire de *direction*. Un simple cas d'arrière-plan serait deux points de contact, séparés dans l'espace et le temps sur le même mode approximativement que des phénomènes visuels phi, et avec pour tâche de juger la direction de la « marche » (qui est logiquement équivalente à la *suite*, mais qui, phénoménologiquement, se présente comme un jugement plus « immédiat »). Prédiction : des effets standard de type phénomène phi, dépendant de l'ISI, avec une plus grande acuité sur les surfaces de haute résolution telles que les bouts des doigts ou les lèvres.

Supposons maintenant que le sujet lève côte à côte ses index gauche et droit et ait le premier point de contact sur un bout de doigt, et le second sur l'autre. Il y aurait une résolution de direction bien plus mauvaise, étant donné que les comparaisons doivent être bilatérales. Ajoutons alors une « aide » visuelle : on laisse le sujet regarder la stimulation du doigt. Mais on lui donne une fausse entrée visuelle : on règle l'appareil de telle manière que la *direction visuelle* qui se trouve en jeu est l'opposé de la direction mise en jeu par la véritable suite de points de contact. Prédiction : les sujets feront en toute confiance des jugements erronés, négligeant ou rejetant la véritable information séquen-

tielle dont ils disposent grâce aux récepteurs cutanés. Si l'effet est très fort, il peut même négliger des jugements unilatéraux voire des jugements sur le même doigt, qui étaient très exacts en l'absence d'entrée visuelle.

2. *Inversions filmiques.* On demande aux sujets de distinguer de courtes « prises » de cinéma ou de caméras vidéo, dont certaines ont été inversées, ou sur lesquelles il y a des coupures séquentielles ou des anomalies. Les cinéastes connaissent des trucs de métier et ont une science extraordinaire qui comprend aussi les effets dûs aux déformations séquentielles introduites dans les plans du film. Parfois ils rafistolent délibérément des scènes avec des plans qui ne servent plus pour créer des effets spéciaux – pour accroître l'angoisse, ou choquer dans des scènes d'horreur, par exemple. Certains événements sont naturellement très fortement ordonnés ; nous avons tous été amusés de voir des films dans lesquels on pouvait voir un plongeur sortir le premier pied d'un éclaboussement dans la piscine, et sauter, lestement et tout sec, sur le plongeur. D'autres événements sont imperceptiblement réversibles – un drapeau qui flotte au vent par exemple – alors que d'autres encore sont intermédiaires ; cela pourrait exiger un effort d'attention soutenu de dire si le filmage d'une balle qui bondit se déroule en avant ou en arrière. Prédiction : les gens ne seront pas doués du tout pour distinguer les inversions dans lesquelles il n'y a aucun parti pris interprétatif – dans lesquels il faut déceler et se rappeler une séquence pure et simple. Par exemple, en gardant à peu près constantes la continuité de mouvement et les disparités de taille et de forme, les sujets seraient bien plus mauvais pour distinguer (pour réidentifier) les séquences qui n'ont aucune interprétation préconçue de la direction, et de les reconnaître par rapport à leurs inversions ou autres transformations. (Les expériences dans lesquelles on essaie de faire la différence entre telle ou telle mélodie en seraient un analogue auditif.)

3. *Écrire sur votre pied.* Cette expérience est censée déranger des jugements reposant sur l'interprétation des « moments d'arrivée » à « disponibilité centrale ». Supposons que vous preniez un stylo bille et que vous imprimiez quelques lettres sur le côté de votre pied nu, sans être capable de voir ce que vous étiez en train de faire. Les signaux qui proviennent des récepteurs cutanés dans votre pied « confirmeraient » que vos actions intentionnelles d'écriture étaient bien accomplies par le stylo dans votre main.

Ajoutez maintenant la vision indirecte, un écran de télévision montrant votre main en train d'écrire sur votre propre pied, mais avec la caméra placée de telle sorte que la pointe du stylo sur le pied soit obscurcie par la main qui tient le stylo. Ces signaux visuels renforceraient la confirmation de l'exécution de vos actions intentionnelles. Mais à présent, insérez un petit retard dans la bande de la télévision (une ou deux images de 33 msec chacun), en sorte que la confirmation visuelle soit toujours retardée d'un tout petit intervalle. Je prédis que les sujets s'habitueraient facilement à cela. (Je l'espère parce que l'étape qui suit est celle qui est intéressante.) Une fois qu'ils se sont habitués, si l'on éliminait soudainement le retard, ils interpréteraient le résultat comme *le fait que le stylo ait l'air de se pencher*, parce que la perception de la trajectoire de sa pointe serait retardée, relativement à leur entrée visuelle, comme si la pointe se déplaçait en prenant conscience de sa trajectoire attendue.

4. *Ajuster le retard sur le carrousel de Grey Walter* : l'expérience qui suit la précédente et mesure le retard requis pour éliminer l'effet de « carrousel précognitif ». Je prédis que la quantité sera bien moindre que les 300-500 msec qui seraient prédites par une extension du modèle stalinien de Libet.

Sur les modèles en Pandémonium du choix de mots

Comment montrer que « les mots veulent se laisser dire » ? Peut-on contrôler expérimentalement l'inspiration ? Les expériences de Levelt ont jusqu'à ce jour produit des résultats négatifs surprenants (voir la note 4, p. 314). Les variantes de ces expériences que j'aimerais voir ouvriraient aux sujets la possibilité de faire un usage « créatif » des mots, tout en fournissant discrètement des matériaux bruts de l'environnement qu'ils puissent incorporer dans leur production. Par exemple, les sujets pourraient être préparés pour l'expérience selon deux dispositifs préalables distincts, dans lesquels on laisserait « traîner » différents mots frappants, éclatants, quelque peu nouveaux ou déplacés (sur des affiches murales, dans les instructions données aux sujets, etc.) ; on leur donnerait alors des possibilités de s'exprimer sur des sujets dans lesquels il serait peu probable que l'on utilise ces expressions, en sorte que l'amorçage du préambule montrerait

que ces expressions cibles ont été « mises en marche » et qu'elles étaient là en attente, prêtes à se laisser utiliser. Si cela ne produisait aucun effet, cela renforcerait le modèle de Levelt ; si cela produisait un effet important (en particulier si des possibilités « tirées par les cheveux » se trouvaient exploitées), cela renforcerait la plausibilité d'un modèle en Pandémonium.

Expérience utilisant des détecteurs de vision

1. *« La vision aveugle » chez les sujets normaux* : les expériences qui utilisent des détecteurs de vision avec des sujets normaux ont montré que quand un stimulus parafovéal est déplacé durant une saccade, les sujets ne s'en aperçoivent pas (ils ne rapportent pas avoir ressenti un déplacement), mais il y a des effets d'amélioration – les latences nécessaires pour l'identification du second stimulus sont plus courtes ou ne dépendent pas de l'information reçue du premier stimulus parafovéal. Si l'on impose aux sujets, dans ces conditions, de faire une conjecture sur le déplacement du stimulus (ou si le premier stimulus était, disons, une lettre majuscule ou minuscule), feront-ils mieux que s'ils s'en remettent au hasard ? Je prédis que oui, pour un ensemble de cas important, mais qu'ils ne font pas mieux que la vision aveugle à son optimum.

2. *Les expériences avec le « papier mural »* : en utilisant un détecteur de vision, et en variant les traits globaux et fins des régions parafovéales de champs répétitifs de « papier mural » durant les saccades, mettez en marche la compétition requise pour rejeter la conclusion selon laquelle il y a « plus de Marilyns ». (Puisque les nouveaux résultats de Ramachandran et de Gregory m'ont ébahi, je me découvrirai et prédirai qu'il n'y a *pas* d'effets *graduels* détectables, bien qu'aux niveaux auxquels les sujets remarquent les changements ils puissent rapporter d'étranges mouvements illusoires.)

3. *Le damier coloré*. Une expérience destinée à montrer combien il y a peu de choses dans « le plein du champ visuel ». On donne aux sujets une tâche d'identification visuelle ou d'interprétation qui requiert des saccades multiples d'une scène mouvante : ils regardent des figures animées noires et blanches sur le fond d'un damier coloré de façon aléatoire. Les cases du damier sont rela-

tivement larges – par exemple, le CRT est divisé en une surface de douze sur dix-huit carrés colorés remplis au hasard de différentes couleurs. (Les couleurs sont choisies au hasard en sorte que la trame n'a pas de signification pour la tache visuelle surimposée sur l'arrière-fond.) Il devrait y avoir des différences de luminescence entre les carrés, en sorte qu'il n'y ait pas d'effet Liebmann, et pour chaque carré on devrait préparer une *couleur alternative isoluminescente* : une couleur qui, si on la déplace avec la couleur qui remplit le carré, ne créerait pas de limites de luminescence radicalement différente sur les bords (c'est destiné à garder en repos les détecteurs de luminescence des bords). Supposez à présent que durant les saccades (détectées par les détecteurs de vision) les couleurs du damier soient déplacées ; ceux qui regardent relèveraient un carré ou plus qui changent de couleur plusieurs fois par seconde. Prédiction : il y aura des conditions dans lesquelles les sujets oublieront complètement le fait que de larges portions de l'« arrière-plan » sont abruptement changées dans leurs couleurs. Pourquoi ? Parce que le système visuel parafovéal est avant tout un système d'alarme, composé de récepteurs conçus pour appeler des saccades quand on note un changement ; un tel système ne s'occuperait pas de suivre à la trace les couleurs insignifiantes entre les fixations, et par conséquent ne retiendrait rien pour pouvoir comparer la nouvelle couleur. (Cela dépend, bien entendu, de « la vitesse du film » dans les régions répondant à la couleur parafovéale ; il peut y avoir une période réfractaire lente qui détruit l'effet que je prédis.)

Bibliographie

- Akins, K.A., 1989, *On Pirhanas, Narcissism and Mental Representation : An Essay on Intentionality and Naturalism*, Dissertation de Ph.D., Département de Philosophie, University of Michigan, Ann Arbor.
- , 1990, « Science and Our Inner Lives : Birds of Prey, Bats and the Common (Featherless) Biped », in M. Bekoff and D. Jamieson éd., *Interpretation and Explanation in the Study of Animal Behavior*, vol. 1, Boulder, CO, Westview, p. 414-427.
- Akins, K.A. and Dennett, D.C., 1986, « Who May I say is Calling ? », *Behavioral and Brain Sciences*, 9, p. 517-518.
- Allman, J., Meizin, F. and Mc Guinness, E.L., 1985, « Direction and Velocity – Specific Responses from beyond the Classical Receptive Fields in the Middle Temporal Visual Area », *Perception*, 14, p. 105-126.
- Allport, A., 1988, « What Concept of Consciousness ? », in Marcel and Bisiach éd., 1988, p. 159-182.
- , « Visual Attention », in M. Posner éd., *Foundations of Cognitive Psychology*, Cambridge, MIT Press, p. 631-682.
- Anderson, J., 1983, *The Architecture of Cognition*, Cambridge, MA, Harvard University Press.
- Anscombe, G.E.M., 1967, *Intention*, Oxford, Blackwell.
- , 1965, « The Intentionality of Sensation » : A Grammatical Feature », in R.J. Butler éd., *Analytical Philosophy*, 2^e série, Oxford, Blackwell, p. 160.
- Anton, G., 1899, « Ueber Selbstwahrnehmung der Herderkrankungen des Gehirn durch den Kranken bei Rindenblindheit und Rindentauheit », *Archiv für Psychiatrie und Nervenkrankheiten*, 32, p. 86-127.
- Arnaud, A., 1641, « Quatrièmes Réponses aux Objections », in Descartes, *Œuvres*, éd. Adam et Tannery, Paris, Vrin-CNRS.

- Baars, B., 1988, *A Cognitive Theory of Consciousness*, Cambridge, Cambridge University Press.
- Bach-y-Rita, P., 1972, *Brain Mechanisms in Sensory Substitution*, New York and London, Academic Press.
- Ballard, D. and Feldman, J., 1982, « Connectionnists Models and their Properties », *Cognitive Science*, 6, p. 205-254.
- Bechtel, W. and Abrahamsen, A., 1991, *Connectionism and the Mind : An Introduction to Parallel Processing in Networks*, Oxford, Blackwell, tr. fr. Joëlle Proust, *Le Connexionnisme et l'esprit*, Paris, La Découverte, 1993.
- Bennett, J., 1965, « Substance, Reality, and Primary Qualities », *American Philosophical Quarterly*, 2, 1-17.
- , 1976, *Linguistic Behavior*, Cambridge, Cambridge University Press.
- Bentham, J., 1789, *Introduction to the Principles of Morals and Legislation*, Londres.
- Bick, P.A. and Kinsbourne, M., 1987, « Auditory Hallucination and Subvocal Speech in Schizophrenic Patients », *American Journal of Psychiatry*, 144, p. 222-225.
- Bieri, P., 1990, Commentary at the Conference « The Phenomenal Mind – How is it Possible and Why is it Necessary ? », Zentrum für Interdisziplinäre Forschung, Bielefeld, Allemagne, mai, 14-17.
- Birnbaum, L. and Collins, G., 1984, « Opportunistic Planning and Freudian Slips », *Proceedings, Cognitive Science Society*, Boulder, CO, p. 124-127.
- Bisiach, E., 1988, « The Haunted Brain and Consciousness », in Marcel and Bisiach, 1988.
- Bisiach, E. and Luzzati, C., 1978, « Unilateral Neglect of Representational Space », *Cortex*, 14, p. 129-133.
- Bisiach, E. et Vallar, G., 1988, « Hemineglect in Humans », in F. Boller and G. Grafman éd., *Handbook of Neuropsychology*, vol. 1, New York Elsevier.
- Bisiach, E., Vallar, G., Perani, D., Papagno, C. and Berti, A., 1986, « Unawareness of Disease Following Lesions of the Right Hemisphere : Anosagnosia for Hemiplegia and Anosagnosia for Hemianopia », *Neuropsychologia*, 14, p. 471-482.
- Blakemore, C., 1976, *Mechanics of the Mind*, Cambridge, Cambridge University Press.
- Block, Ned, 1978, « Troubles with Functionalism », in W. Savage éd., *Perception and Cognition : Issues in the Foundation of Psychology*, Minnesota Studies in the Philosophy of Science, vol. IX, p. 261-326.
- , 1981, « Psychologism and Behaviorism », *Philosophical Review*, 90, p. 5-43.

- , 1990, « Inverted Earth », in J. Tomberlin éd., *Philosophical Perspectives*, 4 : *Action Theory and Philosophy of Mind*, 1990, Atascadero, Cal ; Ridgeview Publishing Co, p. 53-79.
- Boghossian, P.A. and Velleman, J.D., 1989, « Colour as a Secondary quality », *Mind*, 98, p. 81-103.
- , 1991, « Physicalist Theories of Color », *Philosophical Review*, 100, p. 67-106.
- Booth, W., 1988, « Voodoo Science », *Science*, 240, p. 274-277.
- Borges, J.L., 1993, *Fictions*, Gallimard, Paris, et *Œuvres complètes*, I, Gallimard, Paris, Pléiade.
- Borgia, G., 1986, « Sexual Selection in Bowerbirds », *Scientific American*, 254, p. 92-100.
- Braitenberg, V., 1984, *Vehicles : Experiments in synthetic Psychology*, Cambridge MIT Press, A Bradford Book.
- Breitmeyer, B.G., 1984, *Visual Masking*, Oxford, Oxford University Press.
- Broad, C.D., 1925, *Mind and its Place in Nature*, Londres, Routledge and Kegan Paul.
- Brooks, B.A., Yates, J.T., and Coleman, R.D., 1980, « Perception of Images Moving at Saccadic Velocities During Saccades and During Fixation », *Experimental Brain Research*, 40, p. 71-78.
- Byrne, R. and Whiten, A., 1988, *Machiavellian Intelligence : Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*, Oxford Clarendon.
- Calvanio, R., Petrone, P.N. et Levine, D.N., 1987, « Left Visual Spatial Neglect is Both Environment Centered and Body-Centered », *Neurology*, 37, p. 1179-1183.
- Calvin, W., 1983, *The Throwing Madonna : Essays on the Brain*, New York, Mc Graw Hill.
- , 1986, *The River that Flows Uphill : A Journey from the Big Bang to the Big Brain* San Francisco, Sierra Club Books.
- , 1987, « The Brain as a Darwin Machine », *Nature*, 330, p. 33-34.
- , 1989, *The Cerebral symphony : Seashore Reflections on the Structure of Consciousness*, New York, Bantam.
- , 1989b, « A Global Brain Theory », *Science*, 240, p. 33-34.
- Camion, J. Lato, R. and Smith, Y.M. 1983, « Is Blindsight Vision ? » *Behavioral and Brain Sciences*, 6, p. 423-486.
- Camus, A., 1942, *Le Mythe de Sisyphe*, Paris, Gallimard.
- Carruthers, P., 1989, « Brute Experience », *Journal of Philosophy*, 86, p. 258-269.
- Castañeda, C., 1968, *The Teachings of Don Juan : A Yaqui Way of Knowledge*, Berkeley, University of California Press, tr. fr. *Les enseignements d'un sorcier Yaqui*, Paris, Gallimard.

- , H.N., 1967, « Indicators and Quasi Indicators », *American Philosophical Quarterly*, 4, p. 85-100.
- , 1968, « On the Logic of Attributions of Self-Knowledge to Others », *Journal of Philosophy*, 65, p. 439-456.
- Changeux, J.-P. and Danchin, A., 1976, « Selective Stabilization of Developing Synapses as a Mechanism for the the Specification of Neuronal Networks », *Nature*, 264, p. 705-712.
- Changeux, J.-P. and Dehaene, S., 1989, « Neuronal Models of Cognitive Functions », *Cognition*, 33, p. 63-109.
- Cheney, D.L. and Seyfarth, R.M., 1990, *How Monkeys see the World*, Chicago, University of Chicago Press.
- Cherniak, C., 1984, *Minimal Rationality*, Cambridge, Mass., MIT Press, A Bradford Book.
- Churchland, P.M., 1985, « Reduction, Qualia and the direct Inspection of Brain States », *Journal of Philosophy*, 82, p. 8-28.
- , 1990, « Knowing Qualia : a Reply to Jackson », p. 67-76, in Churchland, P.M., *A Neurocomputational Perspective : The Nature of Mind and the Structure of Science*, Cambridge, Mass., MIT Press, A Bradford Book.
- Churchland, P.S., 1981a, « On the Alleged Backwards Referral of Experiences and its Relevance to the Mind-body Problem », *Philosophy of Science*, 48, p. 165-181.
- , 1981, « The Timing of Sensations : Reply to Libet », *Philosophy of Science*, 48, p. 492-497.
- , 1986, *Neurophilosophy : Toward a Unified Science of the Mind/Brain*, Cambridge, Mass., MIT Press, A Bradford Book.
- Clark, R. 1975, *The Life of Bertrand Russell*, London, Weidenfeld and Nicolson.
- Cohen, I.D., Kipnis, D. Kunkle, E.C and Kubzansky, P.E., 1955, « Case Report : Observation of a Person with Congenital Insensitivity to Pain », *Journal of Abnormal and Social Psychology*, 51, p. 33-338.
- Cole, David, 1990, « Functionalism and Inverted Spectra », *Synthese*, 82, p. 207-222.
- Crane, H. and Piantadina, T.P., 1983, « On Seeing Reddish Green and Yellowish Blue », *Science*, 222, p. 1078-1080.
- Crick, F., 1984, « Function of the Thalamic Complex : The Searchlight Hypothesis », *Proceedings of the National Academy of Sciences*, 81, p. 4586-4590.
- Crick, F. and Koch, C., 1990, « Towards a Neurobiological Theory of Consciousness », *Seminars in the Neurosciences*, 2, p. 263-275.
- Damasio, A., Damasio, H. and Van Hoesen, G.W., 1982, « Prosopagnosia : Anatomic Basis and Behavioral Mechanisms, *Neurology*, 32, p. 331-341.

- Darwin, C., 1841, *The Descent of Man, and Selection in Relation to Sex*, 2 vol., Londres, Murray, tr. fr. *Les Origines de l'homme*.
- Davis, W., 1985, *The Serpent and the Rainbow*, New York, Simon and Schuster.
- , 1988, *Passage to Darkness : the Ethnobiology of the Haitian Zombie*, Chapel Hill and London, University of North Carolina Press.
- , « Zombification », *Science*, 240, p. 1715-1716.
- Dawkins, M.S., *Animal Suffering ; the Science of Animal Welfare*, Londres, Chapman & Hall.
- , 1987, « Minding and Mattering », in C. Blackemore et S. Greenfield, eds, *Mindwaves*, Oxford : Basil Blackwell, p. 150-160.
- , « From an Animal's Point of View : Motivation, Fitness and Animal Welfare », *Behavioral and Brain Sciences*, 13, p. 1-61.
- Dawkins, R., 1976, *The Selfish Gene*, Oxford : Oxford University Press.
- , 1991, tr. fr. *Le Gène égoïste*, Paris, A. Colin.
- , 1982, *The Extended Phenotype*, San Francisco, Freeman.
- , 1986, *The Blind Watchmaker*, New York, Norton, tr. fr. *L'Horloger aveugle*, Paris, R. Laffont, 1991.
- De Sousa, R., 1976, « Rational Homunculi », in Amelie O. Rorty éd., *The Identity of Persons*, Berkeley, University of California Press, p. 217-238.
- Dennett, D. C., 1969, *Content and Consciousness*, Londres, Routledge and Kegan Paul.
- , 1971, « Intentional systems », *Journal of Philosophy*, 8, p. 87-106, tr. fr. J. Kahlfa, *Philosophie*, 1, 1984 (repris dans Dennett, 1978a).
- , 1974, « Why the Law of Effect Will not Go Away », *Journal of the Theory of Social Behaviour*, 5, p. 169-197 (repris dans Dennett, 1978 a).
- , 1975, « Are Dreams Experiences ? », *Philosophical Review*.
- , 1978b, « Skinner Skinned », ch. 4 de Dennett 1978a, p. 53-70.
- , 1978c, « Two Approaches to Mental Images », ch. 10 in Dennett, 1978a, p. 174-189.
- , 1978, « Where Am I ? » ch. 17 in Dennett, 1978a, p. 310-323.
- , 1979a, « On the Absence of Phenomenology », in D. Gustafson et B. Tapscott éd., *Body, Mind and Method, Essays in Honor of Virgil Aldrich*, Dordrecht, Reidel, 1979.
- , 1979b, Review of Popper and Eccles, *The Self and its Brain : an Argument for Interactionism*, in *Journal of Philosophy*, 76, p. 91-97.
- , 1981a, « Reflections » on « Software » in Hofstadter et Dennett, 1981.
- , 1981, « Wondering where the Yellow Went » (commentary on W. Sellars's Carus Lectures), *Monist*, 64, p. 102-108.
- , 1982a, « How to Study Consciousness Empirically, or Nothing Comes to Mind », *Synthese*, 59, p. 159-180.

- , 1982b, « Why We Think What We do About Why we Think What We do » : Discussion on Goodman's « On Thoughts without Words », *Cognition*, 12, p. 219-227.
- , 1982c, « Comments on Rorty », *Synthese*, 59, p. 349-356.
- , 1982d, « Notes on Prosthetic Imagination », *New Boston Review*, June, p. 3-7.
- , 1983, « Intentional Systems in Cognitive Ethology : the « Panglossian Paradigm Defended », *Behavioral and Brain Sciences*, 6, 343-390 (repris dans Dennett, 1987).
- , 1984b, « Carving the Mind at its Joints », a Review of Fodor, *The Modularity of Mind*, in *Contemporary Psychology*, 29, p. 284-286.
- , 1985a, « Can Machines Think ? », in M. Shafto éd., *How we Know*, New York, Harper and Row, p. 121-145.
- , 1985b, « Music of the Hemispheres », a review of M. Gazzaniga, *The Social Brain*, in *New York Times Book Review*, November, 17, 1985, p. 53.
- , 1986, « Julian Jaynes Software Archeology », *Canadian Psychology*, 27, p. 149-154.
- , 1987a, *The Intentional Stance*, Cambridge, Mass., MIT Press, Bradford Book, tr. fr. Engel, *La Stratégie de l'interprète*, Paris, Gallimard, 1990.
- , 1987b, « The Geography of Computational Approaches : a View from the East Pole », in R. Harnish and M. Brand éd., *Problems in the Representation of Knowledge*, Tucson, University of Arizona Press.
- , 1988a, « Quining Qualia », in Marcel et Bisiach, 1988, p. 42-77.
- , 1988b, « When Philosophers Encounter AI », *Daedalus*, 117, p. 283-296, réimprimé dans Graubard, 1988.
- , 1988c, « Out of the Armchair and into the Field », *Poetics Today*, 9, numéro spécial sur l'interprétation en contexte dans la science et la culture, p. 205-222.
- , 1988d, « The Intentional Stance in Context and Science and Culture », in Whiten and Byrne, 1988, p. 180-202.
- , 1988e, « Science, Philosophy and Interpretation », *Behavioral and Brain Sciences*, 11, p. 535-546.
- , 1988f, « Why Everyone is a Novelist », *Times Literary Supplement*, Septembre, 16-22.
- , 1989a, « Why Creative Intelligence is Hard to Find », commentaire sur Whiten et Byrne, *Behavioral and Brain Sciences*, 11, p. 253.
- , 1989b, « The Origin of Selves », *Cogito*, 2, p. 163-173.
- , 1989c, « Murmurs in the Cathedral », review of R. Penrose, *The Emperor's New Mind*, in *Times Literary Supplement*, September 29-oct. 5, p. 1066-1068.

- , 1989d, « Cognitive Ethology : Hunting for Bargains or a Wild Goose Chase ? », in A. Montefiore et D. Noble éd., *Own Goals and No goals : a Debate on Goal Directed and Intentional Behaviour*, Londres, Unwin Hyman.
- , 1990a, « Memes and the Exploitation of Imagination », *Journal of Aesthetics and Art Criticism*, 48, p. 127-135.
- , 1990b, « Thinking with a Computer », in H. Barlow éd., *Image and Understanding*, Cambridge, Cambridge University Press, p. 297-309.
- , 1990c, « Betting your Life on an Algorithm », commentary on Penrose, *Behavioral and Brain Sciences*, 13, p. 660.
- , 1990d, « The Interpretation of Texts, People and other Artifacts », *Philosophy and Phenomenological Research*, 50, p. 177-194.
- , 1990e, « Two Black Boxes : A Fable », Tufts University Center for Cognitive Science, Preprint, November.
- , 1991a, « Real Patterns », *Journal of Philosophy*, 89, p. 27-51.
- , 1991b, « Producing Future by Telling Stories », in K.M. Ford and Z. Pylyshyn éd., *Robots Revisited : The Frame Problem in Artificial Intelligence*, Ablex Series in Theoretical Issues in Cognitive Science, Norwood, NJ, Ablex.
- , 1991c, « Mother Nature versus the Walking Encyclopedia », in W. Ramsey, S. Stich, et D. Rumelhart éd., *The Future of Folk Psychology : Intentionality and Cognitive Science*, Cambridge, Cambridge University Press, 1991.
- , 1991e, « Granny's Campaign for Safe Science » in G. Rey and B. Loewer éd., *Fodor and His Critics*, Oxford, Blackwell.
- Dennett, D. and Kinsbourne, M., 1991, « Time and the Observer : The Where and When of Consciousness and in the Brain », *Behavioral and Brain Sciences*.
- Descartes, R., 1637, *Discours de la méthode*, Paris.
- , 1641, *Méditations métaphysiques*, Paris.
- , 1664, *Traité de l'homme*, Paris.
- Dreyfus, H., 1979, *What Computers Can't Do* (2^e éd.), New York, Harper and Row, tr. fr. *Ce que les ordinateurs ne peuvent pas faire*, Seuil.
- Dreyfus, H.L. and S.E., 1988, « Making a Mind Versus Modeling the Brain » : Artificial Intelligence Back at a Branchpoint », in Graubard, 1988.
- Eccles, J.C., 1985, « Mental Summation : The Timing of Voluntary Intentions by Cortical Activity », *Behavioral and Brain Sciences*, 8, p. 542-547.
- Eco, U., 1990, « After Secret Knowledge », *Times Literary Supplement*, June, 22-28, p. 666, « Some Paranoid Readings », *Times Literary Supplement*, June 29-July 5, p. 694.

- Edelman, G., 1987, *Neural Darwinism*, New York, Basic Books.
- , 1989, *The Remembered Present : A Biological Theory of consciousness*, New York, Basic Books, tr. fr. *Biologie de la conscience*, O. Jacob, 1992.
- Efron, R., 1967, « The Duration of the Present », *Proceedings of the New York Academy of Sciences*, 8, p. 542-543.
- Eldredge, N. et Gould, S.J., 1972, « Punctated Equilibria : an Alternative to Phyletic Gradualism », in T.J.M. Schopf éd., *Models of Paleobiology*, San Francisco : Freeman Cooper, p. 82-115.
- Ericsson, K.A. et Simon, H.A., 1984, *Protocol analysis : Verbal Reports as Data*, Cambridge, Mass., MIT Press/A Bradford Book.
- Evans, G., 1982, John Mc Dowell éd., *The Varieties of Reference*, Oxford, Oxford University Press.
- Ewert, J.P., 1987, « The Neuroethology of Releasing Mechanisms : Prey Catching in Toads », *Behavioral and Brain Sciences*, 10, p. 337-405.
- Farah, M.J., 1988, « Is Visual Imagery Really visual ? Overlooked Evidence from Neuropsychology », *Psychological Review*, 95, p. 307-317.
- Farrell, B.A., 1950, « Experience », *Mind*, 59, p. 170-198.
- Fehling, M., Baars, B. et Fischer, C., 1990, « Functional role of Repression in an Autonomous, Resource-Constrained Agent », in *Proceedings of Twelfth Annual conference of the Cognitive Science Society*, Hillsdale, NJ, Erlbaum.
- Fehrer, E. et Raab, D., 1962, « Reaction Time to Stimuli Masked by Metacontrast », *Journal of Experimental Psychology*, 63, p. 143-147.
- Feynmann, R., 1985, *Surely you're Joking, Mr Feynman !*, New York, Norton, tr. fr. *Vous voulez rire M. Feynman*, Paris, InterEditions.
- Finke, R.A., Pinker, S. et Farah, M.J., 1989, « Reinterpreting Visual Patterns in Mental Imagery », *Cognitive Science*, 13, p. 143-147.
- Flanagan, O., 1991, *The Science of the Mind*, Cambridge, Mass., MIT Press/A Bradford Book.
- Flohr, H., 1990, « Brain Processes and Phenomenal Consciousness : A New and Specific Hypothesis », présenté au colloque « The phenomenal Mind – How is it Possible and Why is it Necessary ? », Zentrum für Interdisziplinäre Forschung, Bielefeld, Germany, May 14-17.
- Fodor, J., 1975, *The Language of Thought*, Scranton, PA, Crowell.
- , 1983, *The Modularity of Mind*, Cambridge, Mass., MIT Press/A Bradford Book, tr. fr. A. Gerschenfeld, *La Modularité de l'esprit*, Paris, Minuit, 1986.
- , 1990, *A Theory of Content, and Other Essays*, Cambridge, Mass., MIT Press/A Bradford Book.

- Fodor, J. et Pylyshyn, Z., 1988, « Connexionnism and Cognitive Architecture, A Critical Analysis », *Cognition*, 28, p. 3-71.
- Fox, I., 1989, « On the Nature and cognitive function of Phenomenal Content – Part One », *Philosophical Topics*, 17, p. 91-117.
- French, R., 1991, « Subcognition and the Turing Test », *Mind*, à paraître.
- Freud, S., *Le Moi et le Ça*, Paris, Payot.
- Freyd, J., 1989, « Dynamic Mental Representations », *Psychological Review*, 94, p. 427-438.
- Fuster, J.M., 1981, « Prefrontal Cortex in Motor control », in *Handbook of Physiology, section 1 : the Nervous System, vol. II : Motor control*, American Physiological Society, p. 1149-1178.
- Gardner, H., 1975, *The Shattered Mind*, New York, Knopf.
- Gardner, M., 1981, « The Laffer Curve and Other Laughs in Current Economics », *Scientific American*, 245, December, p. 18-31, repris in Gardner, 1986.
- , 1986, *Knotted Doughnuts and other Mathematical Diversions*, San Francisco, W.H. Freeman.
- Gazzaniga, M., 1978, « Is Seeing Believing : Notes on Clinical Recovery », in S. Finger éd., *Recovery From Brain Damage : Research and Theory*, New York, Plenum Press, p. 409-414.
- , 1985, *The Social Brain : Discovering the Networks of the Mind*, New York, Basic Books.
- Gazzaniga, M. and S. Ledoux, J., 1978, *The Integrated Mind*, New York, Plenum Press.
- Geldard, F.A., 1977, « Cutaneous Stimuli, Vibratory and Saltatory », *Journal of Investigative Dermatology*, 69, p. 83-87.
- Geldard, F. and Sherrick, C.E., 1972, « The Cutaneous “ Rabbit ” : A Perceptual Illusion », *Science*, 178, p. 178-179.
- , 1983, « The Cutaneous Saltatory Area and its Presumed Neural Base », *Perception and Psychophysics*, 33, p. 299-304.
- , « Space, Time and Touch », *Scientific American*, 254, p. 90-95.
- Gert, B., 1965, « Imagination and Verifiability », *Philosophical Studies*, 16, p. 44-47.
- Geshwind, N. et Fusillo, M., 1966, « Color-Naming Defects in Association with Alexia », *Archives of Neurology*, 15, p. 137-146.
- Gide, A., 1948, *Les Faux Monnayeurs*, Paris, Gallimard.
- Goodman, N., 1978, *Ways of Worldmaking*, Hassock, Sussex, Harvester, tr. fr. M.D. Popelard, *Manières de faire des mondes*, J. Chambon, Nîmes, 1991.
- Goody, J., 1977, *The Domestication of the Savage Mind*, Cambridge, Cambridge University Press.
- Gould, S., 1980, *The Panda Thumb*, New York : Norton, tr. fr. *Le pouce du Panda*, Paris, Fayard.

- Gouras, P., 1984, « Color Vision », in N. Osborn and J. Chader éd., *Proccess in Retinal Research*, vol. 3, Londres; Pergamon Press, p. 227-388.
- Graubard, S.R., 1988, *The Artificial Intelligence Debate : False Starts, Real Foundations* (réimpression de *Dedalus*, 117, Hiver 1988), p. 377-388.
- Grey Walter, W., 1963, Presentation to the Osler Society, Oxford University.
- Grice, H.P., 1957, « Meaning », *Philosophical Review*, 66, p. 377-388.
- , 1969 « Utterer's Meaning and Intentions », *Philosophical Review*, 78, p. 147-177.
- Hacking, I., 1990, « Signing », review of Sacks, 1989, *London Review of Books*, April 5, 1990, p. 3-6.
- Hampl, P., 1989, « The Lax Habits of the Free Imagination », *New York Times Book Review*, March 5, 1989, p. 1-39, extrait de Hampl éd., 1989, *The Houghton Mifflin Anthology of Short Fiction*, Boston, Houghton Mifflin.
- Handford, M., 1987, *Where's Waldo ?* Little Brown, Boston.
- Hardin, C.L., 1988, *Colors for Philosophers : Unweaving the Rainbow*, Indianapolis, Hackett.
- , « Color and Illusion », présenté au colloque « The Phenomenal Mind – How is it possible and When is it Necessary ? », Zentrum für Interdisziplinäre Forschung, Bielefeld, Allemagne, May 14-17.
- Harman, G., 1990, « The Intrinsic Quality of Experience », in J.E. Tomberlin éd., *Philosophical Perspectives, 4 : Action Theory and Philosophy of Mind*, Atascadero, CA, Ridgeview, p. 31-52.
- Harnad, S., 1982, « Consciousness : an Afterthought », *Cognition and Brain Theory*, 5, p. 29-47.
- , 1989, « Editorial Commentary », *Behavioral and Brain Sciences*, 12, p. 183.
- Haugeland, J., 1981, *Mind Design : Philosophy, Psychology, Artificial Intelligence*, Montgomey, VT, Bradford Books.
- , 1985, *Artificial Intelligence : The Very Idea*, Cambridge, Mass., MIT Press/A Bradford Book, tr. fr. *L'intelligence artificielle*, Paris, InterEditions.
- Hawkins, S., 1988, *A Brief History of Time*, New York, Bantam.
- Hayes, P., 1979, « The Naive Physics Manifesto », in D. Michiet éd., *Expert Systems in the Microelectronics Age*, Edinburgh University Press.
- Hayes-Roth, B., 1985, « A Blackboard Architecture for control », *Artificial Intelligence*, 26, p. 251-321.
- Hebb, D., 1949, *The Organization of Behaviour : A Neuropsychological Theory*, New York, Wiley.
- Hilbert, D.R., 1987, *Color and color Perception : A Study in Anthro-*

- pocentric Realism*, Sanford University, Center for the Study of Language and Information.
- Hintikka, J., 1962, *Knowledge and Belief*, Ithaca, Cornell University Press.
- Hinton, G.E. and Nowland, S.J., 1987, « How Learning Can Guide Evolution », *Complex Systems, I*, Technical Report, CMU-CS, 86-128, Carnegie Mellon University, p. 495-502.
- Hobbes, T., 1651, *Leviathan*, tr. Tricaud, Sirey, Paris, 1970.
- Hodstadter, D.R. et Dennett, D., 1981, *The Mind's I : Fantasies and Reflections on Self and Soul*, New York, Basic Books, p. 191-201, tr. fr. *Vues de l'esprit*, Paris, InterEditions, 1986.
- Hoffman, R.E., 1986, « What Can Schizophrenic "Voices" Tell Us ? » *Behavioral and Brain Sciences*, p. 535-548.
- Hoffman, R.E. and Kravitz, R.E., 1987, « Feedforward Action Regulation and the Experience of Will », *Behavioral and Brain Sciences*, 10, p. 782-783.
- Hofstadter, D.R., 1981a, « The Turing Test : A Coffeehouse Conversation », in « Metamagical Themas », *Scientific American*, May 1981, repris in Hodstadter et Dennett, 1981.
- , 1981b, « Reflections on Nagel, in Hodstadter et Dennett, 1981, p. 403-414.
- , 1983, « The Architecture of Jumbo », *Proceedings of the Second Machine Learning Workshop*, Monticello, IL.
- , 1985, « On the Seeming Paradox of Mechanizing Creativity », in *Metamagical themas*, New York : Basic Books, p. 526-546, tr. fr. *Mathémagie*, Paris, InterEditions.
- Holland, J.H., 1975, *Adaptation in Natural and Artificial Systems*, Ann Arbor, University of Michigan Press.
- Holland, J.H. et Holyoak, K.J., Nisbett, R.E., et Thagard, P., 1986, *Induction : Processes of Inference, Learning and Discovery*, Cambridge, Mass., MIT Press/A Bradford Book.
- Hondeich, T., 1984, « The Time of Conscious Sensory Experience and Mind-Brain Theories », *Journal of Theoretical Biology*, 110, p. 115-129.
- Howell, R., 1979, « Fictional Objects : How They Are and How they Aren't », in D.F. Gustafson and B. Tapscott éd., *Body, Mind and Method*, Dordrecht, D. Reidel, p. 241-294.
- Hughlins Jackson, J., 1915, « Hughlins Jackson on Aphasia and Kindred Affections of Speech », *Brain*, 38, p. 1-190.
- Hume, D., 1739, *Treatise on Human Nature*, London, John Noon, tr. fr. A. Leroy, Paris, Aubier.
- Humphrey, N., 1972, « Interest » and « Pleasure » : « Two Determinants of a Monkey's Visual Preferences », *Perception*, 1, p. 395-416.
- , 1976, « The Color Currency of Nature », in *Colour for Architecture*,

- T. Porter and B. Mikellides éd., Londres, Studio vista, p. 147-161, repris dans Humphrey, 1983a.
- , 1983a, *Consciousness Regained*, Oxford, Oxford University Press.
- , 1983b, « The Adaptativeness of Mentalism ? », commentary on Dennett, 1983, *Behavioral and Brain Sciences*, 6, p. 366.
- , 1986, *The Inner Eye*, London, Faber and Faber.
- , 1992, *A History of the Mind*, Simon and Schuster, Harvester Wheatsheaf, Hemel Hempstead.
- Humphrey, N. et Dennett, D., 1989, « Speaking for Ourselves : An Assessment of Multiple Personality Disorder », *Raritan*, 9, p. 68-98.
- Humphrey, N. et Keeble, G., 1987, « Effects of Red Light and Loud Noise on the Rates at Which Monkeys Sample the Sensory environment », *Perception*, 7, p. 343.
- Hundert, E., 1987, « Can Neuroscience Contribute to Philosophy », in C. Blackmore and S. Greenfield, *Mindwaves*, Oxford, Basil Blackwell, p. 407-429 (repris comme ch. 7 de Hundert, *Philosophy, Psychiatry and Neuroscience : Three Approaches to the Mind*, Oxford : Clarendon Press 1989).
- Huxley, T., 1874, « On the Hypothesis that Animals are Automata », in *Collected Essays*, London, 1893-1893.
- Jackendof, R., 1987, *Consciousness and the Computational Mind*, Cambridge, Mass., MIT Press/A Bradford Book.
- Jackson, F., 1982, « Epiphenomenal Qualia », *Philosophical Quarterly*, 32, p. 127-136.
- Jacob, F., 1982, *Le Possible et le réel*, Paris, Fayard.
- Janlert, L.E., 1985, *Studies in Knowledge Representation*, Umea, Suède, Institute of Information Processing.
- Jarrell, R., 1963, *The Bat Poet*, New York, Academic Press.
- Jaynes, J., 1976, *The Origins of Consciousness in the Breakdown of the Bicameral Mind*, Boston, Houghton Mifflin.
- Jerison, H., 1973, *Evolution of the Brain and Intelligence*, New York, Academic Press.
- Johnson-Laird, P., 1983, *Mental Models : Towards a cognitive science of Language, inference and consciousness*, Cambridge, Cambridge University Press.
- , 1988, « A Computational Analysis of Consciousness », in A.J. Marcel and E. Bisiach éd., *Consciousness in Contemporary Science*, Oxford, Oxford University Press, New York, Oxford University Press.
- Julesz, B., 1971, *Foundations of Cyclopean Perception*.
- Keller, H., 1808, *The World I live in*, New York, Century Co. (tr. fr.).
- Kinsbourne, M., 1974, « Lateral Interactions in the Brain », in M. Kinsbourne and W.L. Smith éd., *Hemisphere Disconnections and Cerebral Function*, Springfield, Il, Charles C. Thomas, p. 239-259.

- Kinsbourne, M. et Hicks, R.E., 1978, « Functional Cerebral space : A Model for Overflow, Transfer and Inference Effects in Human, Performance : A Tutorial Review », in J. Requin éd., *Attention and Performance*, 7, Hillsdale, NJ, Erlbaum, p. 345-362.
- Kinsbourne, M. et Warrington, E.K., 1963, « Jargon Aphasia », *Neuropsychologia*, 1, p. 27-37.
- Kirman, B.H. et al., 1968, « Congenial Insensitivity of Pain in an Imbecile Boy », *Developmental Medicine and Child Neurology*, 10, p. 57-63.
- Kitcher, Patricia, « Phenomenal Qualities », *American Philosophical Quarterly*, 16, p. 123-129.
- Koestler, Arthur, 1967, *The Ghost in the Machine*, New York, Macmillan.
- Kohler, I., 1961, « Experiments with Goggles », *Scientific American*, 206, p. 62-86.
- Kolers, P.A., *Aspects of Motion Perception*, Londres, Pergamon Press.
- Kolers, P.A. et von Grünau, M., 1976, « Shape and Color in Apparent Motion », *Vision Research*, 16, p. 329-335.
- Kosslyn, S.M., Holtzmann, J.D., Gazzaniga, M.S., et Farrah, M.J., 1985, « A Computational Analysis of Mental Imagery Generation : Evidence for Functional Dissociation in Split Brain Patients », *Journal of Experimental Psychology : General*, 114, p. 311-341.
- Lackner, J.R., « Some Proprioceptive Influences of the Perceptual Representation of Body Shape and Orientation », *Brain*, 11, p. 281-297.
- Langton, C.G., 1989, *Artificial Life*, Redwood City, CA, Addison Wesley.
- Larkin, S. et Simon, H.A., 1987, « Why a Diagram is (Sometimes) Worth Ten Thousand Words », *Cognitive Science*, Oxford, Blackwell.
- Leiber, J., 1988 « "Cartesian" Linguistics ? », *Philosophia*, 118, p. 309-346.
- , 1991, *Invitation to Cognitive Science*, Oxford, Blackwell.
- Leibniz, G.W., 1840, *Monadologie*, éd. Erdmann, in *Opera Philosophica*, Berlin, 2 vol., et éd. Boutroux, Paris, Delagrave.
- Levelt, W., 1989, *Speaking*, Cambridge, Mass., MIT Press/A Bradford Book.
- Levy, J. et Trevarthen, C., 1976, « Metaccontrol of Hemispheric Function in Human Split Brain Patients », *Journal of Experimental Psychology : Human Perception and Performance*, 3, p. 299-311.
- Lewis, D., 1978, « Truth in Fiction », *American Philosophical Quarterly*, 15, p. 37-46.
- , 1979, « Attitudes *de dicto* and *de se* », *Philosophical Review*, 78, p. 513-543.
- , 1988, « What Experience Teaches », *Proceedings of the Russellian*

- Society of the University of Sydney, repris dans W. Lycan éd., *Mind and Cognition, A Reader*, Blackwell, Oxford, 1990.
- Liberman, A. et Studdert-Kennedy, M., 1977, « Phonetic Perception », in R. Held, H. Leibovitz, and H.L. Teuber éd., *Handbook of Sensory Physiology, vol. 8, Perception*, Heidelberg, Springer-Verlag.
- Libet, B., 1965, « Critical Activation in Conscious and Unconscious Experience », *Perspectives in Biology and Medicine*, 9, p. 77-86.
- , 1981, « The Experimental Evidence for Subjective Referral of a Sensory Experience backwards in Time : Reply to P.S. Churchland », *Philosophy of Science*, 48, p. 182-197.
- , 1982, « Brain Stimulation in the Study of Neuronal Functions for Conscious Sensory Experiences », *Human Neurobiology*, 1, p. 235-242.
- , 1985a, « Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action », *Behavioral and Brain Sciences*, 8, p. 529-566.
- , 1985b, « Subjective Antedating of a Sensory Experience and the Mind Brain Theories », *Journal of Theoretical Biology*, 114, p. 563-570.
- , 1987, « Are the Mental Experiences of Will and Self-Control Significant for the Performance of a Voluntary Act ? », *Behavioral and Brain Sciences*, 10, p. 783-786.
- , 1989, « The Timing of a subjective Experience », *Behavioral and Brain Sciences*, 12, p. 183-185.
- Libet, B., Wright, E.W., Feinstein, B. et Pearl, D.K., 1979, « Subjective Referral of the timing for a conscious Sensory Experience », *Brain*, 102, p. 193-224.
- Liebman, S., 1927, « Ueber das Verhalten fahbiger Formen bei Heligkeitsgleichheit von Figur und Grund », *Psychologie Forschung*, 9, p. 200-253.
- Livingstone, M.S. and Hubel, D.H., 1987, « Psychophysical Evidence for Separate Channels for the Perception of form, color, Movement and Deppth », *Journal of Neuroscience*, 7, p. 346-368.
- Lloyd, M. et Dybas, H.S., 1966, « The Periodical Cicada Problem », *Evolution*, 20, p. 132-149.
- Loar, B., 1990, « Phenomenal Properties », in J.E. Tomberlin éd., *Philosophical Perspectives 4, Action Theory and Philosophy of Mind*, Atascadero, Cal., Ridgeview, p. 81-108.
- Locke, J., 1690, *Essay Concerning Human Understanding*, London, Basset, tr. fr. Coste, *Essai sur l'entendement humain*, reprint, Vrin, Paris.
- Lockwood, M., 1989, *Mind, Brain and the Quantum*, Oxford, Blackwell.
- Lodge, D., 1991, *Jeux de société*, Paris, Rivages.
- Lycan, W., 1973, « Inverted Spectrum », *Ratio*, 15, p. 315-319.

- , 1990, « What is the Subjectivity of the Mental ? » in E.J. Tomberlin éd., *Philosophical Perspectives*, 4, *Action Theory and Philosophy of Mind*, Atascadero, Cal., Ridgeview, p. 109-130.
- Marais, E.N., 1937, *The Soul of the White Ant*, London, Methuen.
- Marcel, A.J., 1988, « Phenomenal Experience and Functionalism », in Marcel and Bisiach, 1988, p. 121-158.
- Marcel, A. (sous presse), « Slippage in the Unity of Consciousness », in R. Bornstein et T. Pitmann éd., *Perception without Awareness : Cognitive, Clinical and Social Perspectives*, New York, Guilford Press.
- Marcel, A. et Bisiach, E. éd., 1988, *Consciousness in Contemporary Science*, New York, Oxford University Press.
- Margolis, H.H., 1987, *Patterns, Thinking and Cognition*, Chicago, University of Chicago Press.
- Margulis, L., 1970, *The Origin of Eukaryotic Cells*, New Haven, Yale University Press.
- Marks, C., 1980, *Commissurotomy, Consciousness and Unity of Mind*, Cambridge, Mass., MIT Press/A Bradford Book.
- Marler, P. et Sherman, V., 1983, « Song Structure Without Auditory Feedback : Emendations of the Auditory Template Hypothesis », *Journal of Neuroscience*, 3, p. 517-531.
- Marr, D., 1982, *Vision*, San Francisco, Freeman.
- Maynard-Smith, J., 1978, *The Evolution of Sex*, Cambridge, Cambridge University Press.
- McCulloch, W.S. and Pitts, 1943, « A Logical Calculus for the Ideas Imminent in Nervous Activity », *Bulletin of Mathematical Biophysics*, 5, p. 115-133.
- McGinn, C., 1989, « Can we Solve the Mind-Body Problem ? », *Mind*, 98, p. 349-366.
- , 1990, *The Problem of Consciousness*, Oxford, Blackwell.
- Mc Glynn, S.M. et Schacter, D.L., 1989, « Unawareness of Deficits in Neuropsychological Syndromes », *Journal of Clinical And Experimental Neuropsychology*, 11, p. 143-205.
- Mc Gurk, H. et Mc Donald, R., 1979, « Hearing Lips and Seeing Voices », *Nature*, 264, p. 746-748.
- McClelland, J. et Rumelhart, D. éd., 1986, *Parallel distributed Processing : Explorations in the Microstructure of Cognition*, 2 vols, Cambridge, Mass., MIT Press/A Bradford Book.
- Mc Luhan, M., 1967, *The Medium is the Message*, New York, Bantam.
- Mellor, D.H., 1981, *Real Time*, Cambridge, Cambridge University Press.
- Menzel, E.W., Savage-Rumbaugh, E.S., et Lawson, J., 1985, « Chimpanzee (*Pan Troglodytes*) Spatial Problem Solving with the Use of Mirrors and Televised Equivalents of Mirrors », *Journal of Comparative Psychology*, 99, p. 121-217.

- Millikan, R., 1990, « Truth Rules, Hooverflies and the Kripke-Wittgenstein Paradox », *Philosophical Review*, 99, p. 323-354.
- Minsky, M., 1975, « A framework for Representing Knowledge », Memo, 3306, AI Lab, MIT, Cambridge, Mass., Extraits dans Haugeland, 1981, p. 95-128.
- , 1985, *The Society of Mind*, New York, Simon and Schuster, tr. fr. J. Henry, *La Société des esprits*, Paris, InterEditions, 1990.
- Mishkin, M., Ungerleider, L.G., et Macko, K.A., 1983, « Object Vision and Spatial Vision : Two Cortical Pathways », *Trends in Neuroscience*, 64, p. 370-375.
- Monod, J. *Le Hasard et la nécessité*, Paris, Seuil, 1969.
- Morris, J., K. Rayner, K., et Pollatsek, A., 1990, « Eye Movement guidance in Reading : The Role of Parafoveal and Space Information », *Journal of Experimental Psychology : Human Perception and Performance*, 16, p. 268-281.
- Mountcastle, V.B., 1978, « An Organizing Principle for Cerebral function : The Unit Module and the Distributed system », in G. Edelman et V.B. Mountcastle éd., *The Mindful Brain*, Cambridge, Mass., MIT Press, p. 7-50.
- Nabokov, V., 1930, *Zaschita Luzhina*, in *Sovremennye Zapiski*, Paris, 1930, et sous forme de livre, Berlin, 1930, tr. fr. *La Défense Loujine*, Paris, Gallimard, Folio.
- Nagel, T., 1971, « Brain bisection and the Unity of Consciousness », *Synthese*, 22, p. 396-413, repris dans ses *Mortal questions*, 1979, Cambridge, Cambridge University Press, tr. fr. P. Engel et C. Tiercelin, *Questions mortelles*, Paris, PUF, 1983.
- , 1974, « What is like to be a Bat ? », *Philosophical Review*, 83, p. 435-450, et *Mortal questions*, *ibid.*
- , 1986, *The View from Nowhere*, Oxford, Oxford University Press.
- Neisser, U., 1967, *Cognitive Psychology*, New York, Appleton-Century Croft.
- , 1981, « John Dean's Memory : A Case Study », *Cognition*, 9, p. 435-450.
- , 1988, « Five Kinds of Self-Knowledge », *Philosophical Psychology*, 1, p. 35-39.
- Nemirow, L., 1990, « Physicalism and the Cognitive Role of Acquaintance », in W. Lycan éd., *Mind and Cognition*, A Reader, Oxford, Blackwell, p. 490-499.
- Neumann, O., 1990, « Some Aspects of Phenomenal Consciousness and their Possible Functional Correlates », présenté au colloque « The Phenomenal Mind – How is it Possible and Why is it Necessary ? », *Zentrum für Interdisziplinäre Forschung*, Bielefeld, Allemagne, mai 14-17.
- Newell, A., « Production Systems : Models of Control Structures », in

- W.G. Chase éd., *Visual Information Processing*, New York, Academic Press, p. 463-526.
- , 1982, « The Knowledge Level », *Artificial Intelligence*, 18, p. 81-132.
- , 1988, « The Intentional Stance and the Knowledge Level », *Behavioral and Brain Sciences*, 11, p. 520-522.
- , 1990, *Unified Theories of Cognition*, Cambridge, Mass., Harvard University Press.
- Newell, A., Rosenbloom, P.S. et Laird, J.E., 1989, « Symbolic Architectures for Cognition », in M. Posner éd., *Foundations of Cognitive Science*, Cambridge, Mass., MIT Press, p. 93-132.
- Nielsen, T.I., 1963, « Volition : A New Experimental Approach », *Scandinavian Journal of Psychology*, 4, p. 225-230.
- Nilsson, N., 1984, *Shakey the Computer*, SRI Tech Report, SRI International, Menlo Park, CA.
- Norman, R. and Shallice, T., 1980, *Attention to Action, Willed and Automatic Control of Behavior*, Center for Human Information Processing (Technical Report, n° 99). Réimprimé avec des révisions dans R.J. Davidson, G.E. Schwartz and D. Shapiro éd., *Consciousness and Self Regulation*, New York, Plenum Press.
- Nottelbohm, F., 1984, « Birdsong as a Model in Which to Study Brain Processes Related to Learning », *Condor*, 86, p. 227-236.
- Oakley, D.A. éd., 1985, *Brain and Mind*, Londres et New York, Methuen.
- Ornstein, R. et Thompson, R.F., 1984, *The Amazing Brain*, Boston, Houghton Mifflin.
- Pagels, H., 1988, *The Dreams of Reason : The Computer and the Rise of the Sciences of Complexity*, New York, Simon and Schuster.
- Papert, S., 1988, « One AI or many ? », *Daedalus*, Hiver, p. 1-14.
- Parfit, D., *Reasons and Persons*, Oxford, Clarendon Press.
- Pears, D., 1984, *Motivated Irrationality*, Oxford, Clarendon Press.
- Penfield, W., 1958, *The Excitable Cortex in Conscious Man*, Liverpool, Liverpool University Press.
- Penrose, R., 1989, *The Emperor's New Mind*, Oxford, Oxford University Press ; tr. fr. F. Balibar et C. Tiercelin, *L'esprit, les ordinateurs et les lois de la physique*, Paris, InterEditions, 1991.
- Perlis, 1991, « Intentionality and Defaults », in K.M. Ford and P.J. Hayes éd., *Reasoning Agents in a Dynamic world*, Greenwich, CT, JAI Press.
- Perry, J., 1979, « The Problem of the Essential Indexical », *Nous*, 13, p. 3-21, repris dans son *The Problem of the essential indexical and other essays*, Oxford, Oxford University Press, 1993, article traduit en fr. dans E. Corazza et J. Dokic, *Penser en contexte, la controverse entre G. Evans et J. Perry*, L'Éclat, Combas, 1992.
- Pinker, S. et Bloom, P., 1990, « Natural Language and Natural Selection », *Behavioral and Brain Sciences*, 13, p. 707-784.

- Pollatsek, A., Rayner, K. et Collins, W.E., 1984, « Integrating Pictorial Information Across Eye Movements », *Journal of Experimental Psychology : General*, 113, p. 426-442.
- Pöppel, E., 1985, *Grenzen des Bewusstseins*, Stuttgart : Deutsche Verlags-Anstalt.
- , 1988 (tr. de Pöppel, 1985), *Mindworks : Time and Conscious Experience*, New York, Harcourt Brace Janovich.
- Popper, K.R. et Eccles, J.C., 1977, *The Self and its Brain*, Berlin, Springer Verlag.
- Powers, L., 1978, « Knowledge by Deduction », *Philosophical Review*, 87, p. 337-371.
- Putnam, H., 1965, « Brains and Behaviour », in R.J. Butler éd., *Analytical Philosophy*, Second Series, Oxford, Blackwell, p. 1-19.
- , 1988, « Much Ado about Not Very Much », *Daedalus*, 117, Winter, repris in Graubard, 1988.
- Pylyshyn, Z., 1979, « Do Mental Events Have Durations ? », *Behavioral and Brain Sciences*, 2, p. 277-278.
- Quine, W.V.O., 1969, « Natural Kinds », in *Ontological Relativity and Other Essays*, New York, Columbia University Press, p. 114-138, tr. fr. J. Largeault, *Relativité de l'Ontologie et autres essais*, Paris, Aubier, 1976.
- Ramachandran, V.S., 1985, Guest Editorial in *Perception*, 14, p. 97-103.
- , « 2D or not 2D : That is the Question », in R.L. Gregory, J. Harris, P. Heard, D. Rose, and C. Cronly-Dillon éd., *The Artful Brain*, Oxford, Oxford University Press.
- Ramachandran, V.S. et Gregory, R.L., soumis à *Nature*, « Perceptual Filling in of Artificially Induced Scotomas in Human Vision ».
- Ramsey, W., Stich, S. et Rumelheart, D. éd., 1991, *Philosophy and Connectionist Theory*, Hillsdale, NJ, Erlbaum.
- Raphael, B., 1976, *The Thinking Computer : Mind inside Matter*, San Francisco, Freeman.
- Reddy, D.R., Erman, L.D., Dennel, R.D., et Neely, R.B., « The HEARSAY-II Speech Understanding System : An Example of the Recognition Process », *Proceedings of the International Joint Conference on Artificial Intelligence*, Stanford, p. 185-194.
- Reingold, E.M. et Mericke, P.M., 1990, « On the Interrelatedness of Theory and Measurement in the Study of Unconscious Processes », *Mind and Language*, 5, p. 9-28.
- Reisberg, D. et Chambers, D. (à paraître), « Neither Pictures nor Propositions : What Can we Learn from a Mental Image ? », *Canadian Journal of Psychology*.
- Richards, R.J., 1987, *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, Chicago, University of Chicago Press.

- Ristau, C., 1990, *Cognitive Ethology : The Minds of Other Animals : Essays in Honor of Donald R. Griffin*, Hillsdale, NJ, Erlbaum.
- Rizzolati, G., Gentilucci, M., et Matelli, M., 1985, « Selective Spatial Attention : One Center, One Circuit, or Many Circuits ? », in M.J. Posner and O.S.M. Marin éd., *Attention and Performance XI*, Hillsdale, NJ, Erlbaum.
- Rorty, R., 1970, « Incorrigeability as the Mark of the Mental », *Journal of Philosophy*, 67, p. 399-424.
- , « Comments on Dennett », *Synthese*, 53, p. 323-348.
- Rosenbloom, P.S., Laird, J.E., and Newell, A., « Knowledge-Level Learning in Soar », *Proceedings of AAAI*, Los Altos, CA, Morgan Kaufman.
- Rosenthal, D., 1986, « Two Concepts of Consciousness », *Philosophical Studies*, 49, p. 329-359.
- , 1989, « Thinking that One Thinks », ZIF Report, n° 11, Research Group on Mind and Brain, Perspectives in Theoretical Psychology and the Philosophy of Mind, Zentrum für Interdisziplinäre Forschung, Bielefeld, Allemagne.
- , 1990a « Why Are Verbally Expressed Thoughts Conscious ? » ZIF Report n° 32, Zentrum für Interdisziplinäre Forschung, Bielefeld, Allemagne.
- , 1990b « A theory of Consciousness » ZIF Report, n° 40, Zentrum für Interdisziplinäre Forschung, Bielefeld, Allemagne.
- Rozin, P., 1976, « The Evolution of Intelligence and Access to the cognitive Unconscious », *Progress in Psychobiology and Physiological Psychology*, 6, p. 245-280.
- , « Human Food Selection : The Interaction of Biology, Culture and Individual Experience », in L.M. Barker éd., *The Psychobiology of Human Food Selection*, Westport, CT : Avi Publishing, Co.
- Rozin, P. et Fallon, A.E., 1987, « A Perspective on Disgust », *Psychological Review*, 94, p. 23-47.
- Russell, B., 1927, *The Analysis of Matter*, London, Allen and Unwin, tr. fr. *L'analyse de la matière*, Paris, Payot.
- Ryle, G., 1949, *The Concept of Mind*, London, Hutchinson, tr. fr. S. Stern *Le concept d'esprit*, Paris, Payot.
- , 1979, *On Thinking*, éd. K. Kolenda, Totowa, NJ, Rowman and Littlefield.
- Sacks, O., 1985, *The Man who Mistook his Wife for His Hat*, New York, Summit Books.
- , *Seeing Voices*, Berkeley, University of California Press.
- Sandevall, E., 1991, « Toward a Logic of dynamic Frames », in K.M. Ford and J. Hayes éd., *Reasoning in Agents in a Dynamic World*, Greenwich, CT, JAI Press.

- Sanford, D., 1975, « Infinity and Vagueness », *Philosophical Review*, 84, p. 520-535.
- Sartre, J.P., 1943, *L'Être et le néant*, Paris, Gallimard.
- Schank, R., 1991, *Tell me a Story*, New York, Scribener.
- Schank, R., et Abelson, R., 1977, *Scripts, Plans, goals and Understanding : An Inquiry into Human Knowledge Structures*, Hillsdale, NJ, Erlbaum.
- Schull, J., 1990, « Are Species Intelligent ? », *Behavioral and Brain Sciences*, 3, p. 417-458.
- Searle, J., 1980, « Minds, Brains and Programs », *Behavioral and Brain Sciences*, 3, p. 417-458 (tr. fr. in Hofstadter and Dennett, 1981).
- , 1982, « The Myth of the Computer : an Exchange », *New York Review of Books*, June 24, p. 56-57.
- , 1983, *Intentionality : An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press, tr. fr. C. Pichevin, *L'intentionnalité*, Paris, Minuit, 1987.
- , 1984, « Panel Discussion : Has Artificial Intelligence Research Illuminated Human Thinking ? » éd., *Computer culture : The Scientific, Intellectual and Social Import of the Computer*, Annals of the New York Academy of Sciences, 426.
- , 1988a, « Turing the Chinese Room », in Singh éd., *Synthesis of Science and Religion, Critical Essays and Dialogues*, San Francisco, Bhaktivedanta Institute, 1988.
- , 1988b, « The Realistic Stance », *Behavioral and Brain Sciences*, 11, p. 527-529.
- , 1990a, « Consciousness, Explanatory Inversion and Cognitive Science », *Behavioral and Brain Sciences*, 13, p. 585-642.
- , 1990b, « Is the Mind a Computer Program ? », *Scientific American*, 262, p. 26-31.
- Selfridge, O., 1959 « Pandemonium : A Paradigm for Learning », *Symposium on the Mechanization of Thought Processes*, Londres, HM Stationery Office.
- , *Tracking and Trailing*, inédit.
- Sellars, W., 1963, « Empiricism and the Philosophy of Mind », in *Science, Perception and Reality*, London, Routledge and Kegan Paul, tr. fr. F. Cayla, *Empirisme et Philosophie de l'esprit*, L'Éclat, Combas, 1992.
- Shallice, T., 1972, « Dual Fonctions of Consciousness », *Psychological Review*, 79, p. 383-393.
- , 1978, « The Dominant Action System : An Information Processing Approach to Consciousness », in K.S. Pope et J. L. Singer éd., *The Stream of Consciousness*, New York, Plenum.
- , 1988, *From Neuropsychology of Mental Structure*, Cambridge : Cambridge University Press.

- Sharpe, T., 1977, *The Great Pursuit*, London, Secker and Warburg.
- Shepard, R.N. and Cooper, L.A., 1982, *Mental Images and Their Transformations*, Cambridge, Mass., MIT Press/A Bradford Book.
- Shepard, R.N. et Meltzer, J., 1971, « Mental Rotation of Three Dimensional Objects », *Science*, 171, p. 701-703.
- Shoemaker, S., 1969, « Time without Change », *Journal of Philosophy*, 66, p. 363-381.
- , 1975, « Functionalism and Qualia », *Synthese*, 27, p. 291-315.
- , 1981, « Absent Qualia are Impossible – A Reply to Block », *Philosophical Review*, 90, p. 581-599.
- , 1988, « Qualia and Consciousness », Tufts University Philosophy Department Colloquium.
- Siegel, R.K. et West, L.J. éd., 1975, *Hallucinations, Behavior, Experience and Theory*, New York, Wiley.
- Simon, H.A. et Kaplan, C.A., 1989, « Foundations of Cognitive Science », in Posner éd., *Foundations of Cognitive Science*, Cambridge, Mass., MIT Press.
- Smolensky, P., 1988, « On the Proper Treatment of Connexionism », *Behavioral and Brain Sciences*, 11, p. 1-74.
- Smullyan, R.M., 1981, « An Epistemological Nightmare », in Hostdter and Dennett, 1981, p. 415-427, réimprimé dans Smullyan, 1982, *Philosophical Fantasies*, New York, St Martin's Press.
- Smythies, J.R., 1954, « Analysis of Projection », *British Journal of Philosophy of Science*, 5, p. 120-133.
- Snyder, D.M., 1988, « On the Time of a Conscious Peripheral Sensation », *Journal of Theoretical Biology*, 130, p. 253-254.
- Sperber, D. et Wilson, D., 1986, *Relevance, Communication and Cognition*, Oxford, Blackwell, tr. fr. par D. Sperber et A. Gershenfeld, *La Pertinence, communication et cognition*, Paris, Minuit, 1990.
- Sperling, G., « The Information Available in Brief Visual Presentations », *Psychological Monographs*, 74, n° 11.
- Sperry, R.W., 1977 « Forebrain Commissurotomy and conscious Awareness », *The Journal of Medicine and Philosophy*, 2, p. 101-126.
- Spillman, L. et Werner, J.S., 1990, *Visual Perception : The Neuro-Psychological Foundations*, San Diego, Academic Press.
- Spinoza, B., 1677, *Essai sur la réforme de l'entendement*, tr. Caillois, Paris, Pléiade, Gallimard.
- Stafford, S.P., 1983, « On the Origin of the Intentional Stance », Tufts University Working Papers in cognitive Science, CCM, 83-1.
- Stalnaker, R., 1984, *Inquiry*, Cambridge, Mass., MIT Press/A Bradford Book.
- Stix, G., 1991, « Reach Out », *Scientific American*, 264, p. 134.

- Stoerig, P. and Cowey, A., 1990, « Wavelength Sensitivity in Blind-sight », *Nature*, 342, p. 916-918.
- Stoll, C., 1989, *The Cuckoo's Egg : Tracking a Spy Through the Maze of Computer Espionage*, New York, Doubleday.
- Straight, H.S., 1976, « Comprehension vs Production in Linguistic Theory », *Foundations of Language*, 14, p. 525-540.
- Stratton, G.M., 1896, « Some Preliminary Experiments on Vision without Inversion of the Retinal Image », *Psychology Review*, 3, p. 611-617.
- Strawson, G., 1989, « Red and "Red" », *Synthese*, 78, p. 193-232.
- Strawson, P.F., 1962, « Freedom and Resentment », *Proceedings of the British Academy*, repris dans P.F. Strason éd., *Studies in the Philosophy of Thought and Action*, Oxford, Oxford University Press, 1968.
- Taylor, D.M., 1966, « The Incommunicability of Content », *Mind*, 75, p. 527-541.
- Thomson, d'Arcy, W., 1917, *On Growth and Form*, Cambridge, Cambridge University Press.
- Thomson, E., Palacios, A., et Varela, F. (sous presse), « Ways of Coloring », *Behavioral and Brain Sciences*.
- Tranel, D. et Damasio, A.R., 1988, « Non Conscious Face Recognition in Patients with Face Agnosia », *Behavioral and Brain Research*, 30, p. 235-249.
- Tranel, D., Damasio, A.R. et Damasio, H., 1988, « Intact Recognition of Facial Expression, Gender and Age in Patients with Impaired Recognition of Face Identity », *Neurology*, 38, p. 690-696.
- Treisman, A., 1988, « Features of Objects : the Fourteenth Bartlett Memorial Lecture », *Quarterly Journal of Experimental Psychology*, 40A, p. 201-237.
- Treisman, A. et Gelade, G., 1980, « A Feature-integration theory of Attention », *Cognitive Psychology*, 12, p. 97-136.
- Treisman et Sato, S., 1990, « Conjunction Search, Revisited », *Journal of Experimental Psychology : Human Perception and Performance*, 16, p. 459-478.
- Treisman, A. et Souther, J., 1985, « Search Asymetry : A diagnostic for Preattentive Processing of Separable Features », *Journal of Experimental Psychology : General*, 114, p. 285-310.
- Turing, A., 1950, « Computing Machinery and Intelligence », *Mind*, 59, p. 433-460, tr. fr. G. Gièze, in A.R. Anderson éd., *Pensée et machine*, Champ Vallon, Lyon, 1984.
- Tye, M., 1986, « The Subjective Qualities of Experience », *Mind*, 95, p. 1-17.
- Uttal, W.R., 1979, « Do Central Non Linearities Exist ? » *Behavioral and Brain Sciences*, 2, p. 286.

- Van der Waals, H.G. et Roelofs, C.O., 1930, « Optische Scheinbewegung », *Zeitschrift für Psychologie und Physiologie des Sinnerorgane*, 114, p. 241-288, 115 (1931), p. 91-190.
- Van Essen, D.C., 1979, « Visual Areas of the Mammalian Cerebral cortex », *Annual Review of Neuroscience*, 2, p. 227-263.
- Van Gulick, R., 1988, « Consciousness, Intrinsic Intentionality and Self understanding Machines », in Marcel and Bisiach 1988, p. 78-100.
- , 1989, « What Difference does Consciousness Make ? » *Philosophical Topics*, 17, p. 211-230.
- , 1990, « Understanding the Phenomenal Mind : Are we all just Armadillos ? », présenté au colloque « The Phenomenal Mind – How is it Possible and Why is it Necessary ? », Zentrum für Interdisziplinäre Forschung, Bielfeld, Allemagne, mai 14-17.
- Van Tuijl, H.F.J.M., 1975, « A New Visual Illusion : Neonlike Color Spreading and Complementary Color Induction between Subjective Contours », *Acta Psychologica*, 39, p. 441-445.
- Vendler, Z., 1972, *Res cogitans*, Ithaca, Cornell University Press.
- , 1984, *The Matter of Minds*, Oxford, Oxford University Press.
- Von der Malsburg, C., 1985, « Nervous Structures with Dynamical Links », *Berichte der Bunsen-Gesellschaft für Physicalische Chemie*, 89, p. 703-710.
- Von Uexküll, J., 1909, *Umwelt und Innenwelt der Tiere*, Berlin, Jena, tr. fr. *Mondes animaux et mondes humains*, Paris, Denoël, Conthier, 1964.
- Vosberg, R., Fraser, N. et Guehl, J., 1960, « Imagery Sequence in Sensory Deprivation », *Archives of General Psychiatry*, 2, p. 356-357.
- Walton, K., 1973, « Pictures and Make Believe », *Philosophical Review*, 82, p. 283-319.
- , 1978, « Fearing fiction », *Journal of Philosophy*, 75, p. 6-27.
- Warren, R.M., 1970, « Perceptual Restoration of Missing Speech Sounds », *Science*, 167, p. 392-393.
- Wasserman, G.S., 1985, « Neural/Mental Chronometry of Chronotheology », *Behavioral and Brain Sciences*, 8, p. 556-557.
- Weizkranz, L., 1988a, *Blindsight : A Case Study and Implications*, Oxford, Oxford University Press.
- , 1988b, « Some contributions to Neuropsychology of vision and Memory to the Problem of consciousness », in Marcel and Bisiach, 1988, p. 183-199.
- , 1989, Panel discussion on consciousness, European Brain and Behavior Society, Turn, septembre 1989.
- , 1990, « Outlooks for Blindsight : Explicit Methodologies for Implicit Processes » (the Ferrier Lecture), *Proceedings of the Royal Society, London*, B 239, p. 247-278.

- Welch, R.B., 1978, *Perceptual Modification : Adapting to Altered Sensory Environments*, New York, Academic Press.
- Wertheimer, M., 1912, « Experimentelle Studien über das Sehen von Bewegung », *Zeitschrift für Psychologie*, 61, p. 161-265.
- White, S., 1986, « The Curse of the Qualia », *Synthese*, 68, p. 33-368.
- Whiten, A. et Byrne, R., 1988, « Toward the Next Generation in Data Quality : A New Survey of Primate Tactical Deception », *Behavioral and Brain Sciences*, 11, p. 267-273.
- Wiener, N., 1948, *Cybernetics : or Control and Communication in the Animal and the Machine*, Cambridge, Technology Press, repr. Paris, Hermann, 1964.
- Wilkes, K., *Real People*, Oxford, Oxford University Press.
- Willson, L., 1974, « Observations and Experiments in the Ethology of the European Beaver », *Viltrevy, Swedish Wildlife*, 8, p. 115-266.
- Winograd, T., 1972, *Understanding Natural Language*, New York, Academic Press.
- Wittgenstein, L., 1953, *Philosophische Untersuchungen*, Oxford, Blackwell, tr. fr. P. Klossowski, *Investigations philosophiques*, Paris, Galilimard, 1961.
- Wolfe, J.M., 1990, « Three Aspects of the Parallel Guidance of Visual Attention », *Proceedings of the Cognitive Science Society*, Hillsdale, NJ, Erlbaum, p. 1048-1049.
- Yonas, A., 1981, « Infants Response to Optical Information for Collision », in R.N. Aslin, J.R. Alberts, et M.R. Peterson éd., *Development of Perception : Psychobiological Perspectives*, vol. 2, *The Visual System*, New York, Academic Press.
- Young, J.Z., 1965a, « The Organization of the Memory System », *Proceedings of Royal Society, London, (Biology)*, 163, p. 285-320.
- , 1965b, *A Model of the Brain*, Oxford, Clarendon.
- , 1979, « Learning as a Process of Selection », *Journal of the Royal society of Medicine*, 72, p. 801-804.
- Zajonc R., Marcus, H., 1984, « Affect and Cognition : the Hard Interface », in C. Izard, J. Kagan, and R. Zajonc éd., *Emotion, Cognition and Behavior*, Cambridge University Press, p. 73-102.
- Zeki, S.M. et Shipp, S., 1988, « The Functional Logic of Cortical connexions », *Nature*, 335, p. 311-317.
- Zihl, J., 1980, « Blindsight : Improvement of Visually Guided Eye Movements by systematic Practice in Patients with Cerebral Blindness », *Neuropsychologica*, 18, p. 71-77.
- , 1981, « Recovery of Visual Functions in Patients with Cerebral Blindness », *Experimental Brain Research*, 44, p. 159-169.

Index des noms

- Abelson R. 323, 542.
Abrahamson A. 336.
Akins K. 12, 13, 226, 315, 465, 473.
Allison M. 310.
Allman J., Meizin F. et Mc Guinness E. L. 173.
Allport A. 325, 344, 354.
Anderson J. 13, 323, 331, 332, 334.
Andler D. 12.
Anscombe G. E. M. 399, 414.
Anton G. 105, 442.
Antony M. 13.
Aristote 283.
Arnauld A. 401.
- Baars B. 322, 324, 336, 338, 339, 342, 347, 348, 427, 428, 441.
Bach-y-Rita P. 12, 421, 423, 424, 455.
Ballard D. 238.
Baudot J. 455.
Beattie J. 13.
Bechtel W. 336.
Bennett J. 243, 470.
Bentham J. 557.
Berkeley G. 76, 91.
Bick P. 351.
Bick P. A. et Kinsbourne M. 315.
Bieri P. 12, 351.
Birnbaum L. 303-304.
Bisiach E. 12, 322, 442.
Block N. 13, 178, 494, 540, 545, 565.
Bloom P. 283.
Boghossian P. A. 572.
Booth W. 131.
Borges J. 283, 509.
Borgia G. 515.
- Boyle R. 459.
Braitenberg V. 217, 225, 226.
Breitmeyer B. 181.
Broad C. D. 498.
Bronowsky J. 261.
Brooks B. A. 447.
Byrne R. 283, 553.
- Calvanio R. 442.
Calvin W. 12, 149, 186, 206, 239, 280, 282, 323, 343.
Campion J. 454.
Camus A. 35-37.
Canli T. 13.
Carnegie D. 376.
Carruthers P. 409.
Casper le Gentil Fantôme 53.
Castañeda C. 19, 534.
Castañeda H. N. 534.
Catherine la Grande 22.
Chambers D. 362, 368.
Changeux J.-P. 282.
Chaplin Greta et Freda 523.
Cheney D. 243, 551, 553.
Cherniak C. 354.
Chomsky N. 290, 375.
Churchland P. M. 494.
Churchland P. S. 56, 197, 202-205.
Clark R. 307.
Cohen I. 83.
Colasanti L. 10.
Cole D. 487, 488, 509.
Collins G. 302-304.
Collins W. 448.
Conan Doyle A. 109, 505, 506.
Cooper L. 125, 456.

- Cowey A. 406.
 Crane H. 131.
 Crick F. 210, 342.
 Crick F. 320, 321.

 Dahlbom B. 12, 13.
 Damasio A. 131.
 Damasio H. 131.
 Danchin A. 282.
 Darwin C. 218, 261, 282, 522.
 Davis W. 131.
 Dawkins M. 556, 557.
 Dawkins R. 219, 244, 251-253, 255, 257, 259, 281, 515, 522, 551.
 De Kooning, W. 476.
 De Sousa R. 527.
 Dehaene S. 282.
 Dennett D. 9, 10, 57, 60, 63, 103, 116, 127, 131, 176-178, 185, 197, 215, 220, 221, 230, 232, 240, 242, 280, 281, 283, 308, 314, 315, 327, 337, 349, 354, 363, 387, 395, 398, 445, 457, 460, 469, 491, 504, 509, 512, 526, 532, 545, 553, 565, 572, 573.
 Derrida J. 508.
 Descartes R. 11, 15, 18, 20, 21, 23, 46, 50-52, 59, 60, 88, 91, 92, 95, 97, 138-141, 167, 210, 214, 323, 352, 384, 402, 410, 554, 557.
 Dorfman A. 151, 170.
 Dreyfus H. 338, 348, 349.
 Dreyfus S. 338.

 Eccles J. C. 46, 60, 197, 198, 202, 216.
 Eco U. 306.
 Edelman G. 282, 336, 353, 399, 439, 572.
 Efron R. 189, 216.
 Eldredge N. 522.
 Ericsson 101.
 Evans G. 323.
 Ewert J. P. 236.

 Fallon A. 514.
 Faraday M. 62.
 Farah M. 125, 398.
 Faucher L. 13.
 Fehrer E. 181.
 Feldman J. 238.
 Feynman R. 366, 367.
 Finke R. 398.
 Flanagan O. 13, 341.

 Flohr H. 12, 342.
 Fodor J. 282, 301, 325-327, 349, 353, 354, 377.
 Forster, E. M. 243, 306.
 Fourier 67.
 Fox I. 457, 498.
 Fraser, N. 26.
 French R. 387.
 Freud S. 27, 113, 303, 391.
 Freyd J. 398.
 Fusillo M. 442.
 Fuster J. M. 343.

 Gardner H. 454.
 Gardner M. 145.
 Gazzaniga M. 125, 247, 325, 526.
 Gelade G. 455.
 Geldard F. 183.
 Gentilucci M. 237.
 Gert B. 484.
 Geshwind N. 442.
 Gide A. 207.
 Gilbert W. 62.
 Goodman N. 150, 156, 164, 166, 169.
 Goody J. 324.
 Gould S. 283.
 Gouras P. 465.
 Graubard S. 336.
 Gregory R. 456, 579.
 Grey Walter W. 211, 212.
 Grice H. 283.
 Griffin D. 553.
 Guehl J. 26.

 Hacking I. 556.
 Hampl P. 305, 306.
 Handford M. 347.
 Hardin C. 60, 131, 248, 456, 465, 509, 547.
 Harman G. 457, 494, 498.
 Harnad S. 204, 206, 216, 276, 385, 388.
 Haugeland J. 336.
 Hayes P. 323.
 Hayes-Roth B. 331.
 Hebb D. 261, 282, 571.
 Heesch W. 315.
 Hilbert D. R. 13, 465.
 Hintikka J. 569.
 Hinton J. 233.
 Hobbes T. 167, 178.
 Hoffman R. 315.
 Hofstadter D. 13, 276, 300, 385, 387, 509, 547, 565.

- Holcomb P. 13.
 Holland J. 282.
 Holmes Sherlock 106, 107, 109, 505, 506.
 Holtzman N. 125.
 Holyoak K. 282.
 Honderich T. 197, 200.
 Howell R. 106.
 Hu-Min Ji 565.
 Hubel D. 58.
 Hughlins Jackson H. 243.
 Hume D. 60, 76, 92, 171, 230, 251, 511, 512.
 Humphrey N. 12, 49, 224, 276, 467, 476, 477, 508, 521, 526, 531.
 Hundert E. 427.
 Husserl E. 62.
 Huxley T. 497, 500-502.
- Jackendoff R. 12, 169, 348, 354, 398, 427, 429.
 Jackson F. 493, 540.
 Jacob F. 223.
 Jacob P. 12.
 Jaikumar M. 13.
 James W. 135, 267, 275, 286, 512.
 Janlert R. 366.
 Jarrell R. 550.
 Jaynes J. 276, 324.
 Jerison H. 324.
 Johnson D. 509.
 Johnson-Laird P. 320, 321, 529.
 Joslin D. 13.
 Joyce J. 264.
 Julesz B. 146, 320.
- Kafka F. 532.
 Kant E. 62, 178, 179.
 Keeble G. 476.
 Keller H. 285.
 Kessel F. 509.
 Keynes J. 371, 378.
 Kim J. 12.
 Kinsbourne M. 12, 210, 215, 229, 310, 315, 339, 526.
 Kirman B. H. 83.
 Kissinger H. 110.
 Kitcher P. 457.
 Koch C. 320, 321.
 Koestler A. 60.
 Kohler I. 486.
 Kolers P. 150, 156, 158, 165, 166, 180.
 Kosslyn S. 125, 324, 358, 367, 398, 399.
- Lackner J. 131.
 Laffer A. 143-145.
 Laird J. 331, 334, 337, 353.
 Land E. 59.
 Langton C. 222.
 Larkin J. 336.
 Latto R. 454.
 Lawlor K. 13.
 Ledoux J. 325, 526.
 Leiber J. 88, 376, 377, 547.
 Leibniz G. 511.
 Levelt W. 12, 291, 293-297, 301, 307, 315, 375, 578, 579.
 Levine D.N. 442.
 Levy J. 526.
 Lewis D. 106, 484, 534.
 Liberman A. 70.
 Libet B. 195-209, 211, 212, 216, 427.
 Lincoln A. 304, 305, 307.
 Livingstone M. 173.
 Llinas R. 282.
 Loar B. 494.
 Locke J. 76, 92, 251, 459-462, 464, 465, 482.
 Lockwood M. 460, 509.
 Lodge D. 508.
 Lorentz 177.
 Luzzatti C. 442.
 Lycan W. 484, 494.
- Macko K. 455.
 Marceau M. 263, 269.
 Marcel A. 12, 309, 322, 409, 411.
 Margolis H. 241, 275, 279, 280, 346, 428.
 Margulis L. 513.
 Markus H. 498.
 Marler P. 231.
 Marr D. 344, 345.
 Marx K. 285.
 Matelli M. 237.
 Maxwell C. 62, 177.
 Maynard-Smith J. 281.
 Mc Clelland J. 275, 300.
 Mc Connell J. 13.
 Mc Culloch W. 283.
 Mc Donald R. 147.
 Mc Ginn C. 88, 341, 454, 538.
 Mc Glynn S. 439, 442, 456.
 Mc Gurk H. 147.
 Mc Kay 216.
 Mc Luhan M. 476.
 Mellor H. 190, 193.
 Menzel E. 531.
 Merikle P. 162.

- Miller G. 228.
 Minsky M. 12, 89, 142, 300, 302, 323, 327, 329, 341, 398, 399, 440, 444, 445, 529, 542.
 Mishkin M. 455.
 Monod J. 219.
 Monroe M. 437, 438.
 Morris J. 448.
 Mountcastle V. 328.
 Mynihan D. 441.

 Nabokov V. 372.
 Nagel T. 97, 128, 341, 461, 526, 534, 538, 546-548, 555, 568, 572.
 Neisser U. 25, 192, 215, 349.
 Nemirov L. 494.
 Neumann O. 12, 210, 227, 276.
 Newell A. 215, 320, 331, 334, 337, 353.
 Nielsen T. 177.
 Nietzsche F. 285.
 Nilsson N. 12, 115.
 Nisbett R. 282.
 Norman D. 238, 343, 346.
 Nottelbohm F. 231.
 Nowlan S. 233.

 Oakley K. 526.
 Ornstein R. 458.
 Orwell G. 152.
 Otto 289-290, 302, 378, 379, 386, 394, 397, 414, 419, 444, 448-453, 473, 474, 478-482, 498, 567.

 Pagels H. 206.
 Palacios A. 456, 465, 467, 508.
 Papert S. 283.
 Parfit D. 525.
 Partee, B. 279.
 Pears D. 313.
 Penfield W. 89, 196.
 Penrose R. 54, 198, 533.
 Père Noël 114, 462, 463.
 Perlis D. 347, 529.
 Perry J. 534.
 Petrone P. 442.
 Piantanida T. 131.
 Pinal P. 177.
 Pinker S. 283.
 Pitts W. 283.
 Platon 257, 277, 280, 333, 338, 353.

 Pollatsek A. 448, 455.
 Pöppel E. 140, 142.
 Popper K. 46, 60, 198, 202.
 Potemkine 22.
 Potter B. 528.
 Proust J. 10.
 Putnam H. 349, 484, 572.
 Pylyshyn Z. 215, 353.

 Quine W. 32, 508, 509.

 Ramachandran V. 325, 354, 456, 579.
 Ramberg B. 13.
 Ramset W. 336.
 Raphaël B. 115.
 Rawdon-Hasting F. 215.
 Rayner K. 448, 455.
 Reingold E. M. 162.
 Reisberg D. 362, 368.
 Richard R. 232, 282, 522.
 Ristau C. 553.
 Rizzolati G. 237.
 Roelofs 165.
 Roepke M. 13.
 Rorty R. 93, 572.
 Rosenberg E. 466, 473, 496.
 Rosenberg J. 12, 178, 466, 473, 496.
 Rosenbloom P. 215, 331, 334, 337, 353.
 Rosenthal D. 12, 380, 382-385, 388, 390, 391, 394, 414.
 Rozin P. 324, 476, 514.
 Rumelhart D. 275, 300, 336.
 Russell B. 178, 307, 327.
 Ryle G. 50, 68, 278, 512.

 Sacks O. 374, 454, 556.
 Sandeval E. 323.
 Sanford D. 522.
 Sangree M. 13.
 Sartre J. P. 207.
 Sato S. 455.
 Savage-Rumbaugh S. 509.
 Schacter D. 439, 442, 456.
 Schank R. 323, 542.
 Scheerer E. 12.
 Schlossberger C. 13.
 Schull J. 232.
 Searle J. 128, 338, 348, 349, 402, 540, 542, 543, 545, 565, 568, 572.
 Selfridge O. 12, 237, 238, 415.
 Sellars W. 87, 460, 475.

- Seyfarth R. 243, 551, 553.
 Shakespeare W. 108.
 Shakey 12, 115-117, 119-126, 129,
 168, 282, 288-290, 358, 363, 370,
 386, 388.
 Shallice T. 238, 320, 343, 346, 454.
 Sharpe T. 512, 514.
 Shepard R. 125, 357-362, 364, 370,
 456.
 Sherman V. 231.
 Sherrick C. 183.
 Shoemaker S. 457, 475, 484.
 Shylock 502.
 Siegel R. 26.
 Simon H. 101, 366.
 Skinner B. 231.
 Sloman A. 12.
 Smith Y. 454, 526.
 Smolensky P. 337.
 Smullyan R. 336.
 Smythies J. 167-169, 178.
 Snyder D. 178.
 Souther J. 455.
 Spalding O. 13.
 Sperber D. 12, 244, 283, 314.
 Sperling G. 215.
 Sperry R. 348.
 Spillman L. 131.
 Spinoza B. 236.
 Stafford S. 13, 49.
 Stalnaker R. 354.
 Statton C. 486.
 Steinberg L. 13.
 Stich S. 336.
 Stoerig P. 406.
 Stoll C. 164, 170.
 Straight H. S. 371.
 Strawson G. 487.
 Strawson P. 48.
 Studdert-Kennedy M. 70.
 Sullivan A. 258.
 Superman 40.
 Taylor D. 484.
 Thagard P. 282.
 Thompson d'Arcy. 217.
 Thompson E. 13, 456, 458, 465, 467,
 508.
 Tiercelin C. 10.
 Titchener W. 62.
 Tranel D. 131.
 Treisman A. 455.
 Trevarthen C. 526.
 Truman H. 49, 532.
 Turing A. 262-263, 265, 267-269, 271,
 295, 330, 332, 387, 388, 540, 541,
 545, 570.
 Tye M. 494.
 Ungerleider L. 455.
 Uttal W. 215.
 Valéry P. 524.
 Vallar G. 442.
 Van der Waals H. 165.
 Van Essen D. 173.
 Van Hoesen G. 131.
 Van Gulick R. 12, 348, 349, 409, 494.
 Van Tuijl. H. 434.
 Varela F. 12, 465, 467, 508.
 Vendler Z. 60.
 Vesale 138.
 Von der Heiden L. 12.
 Von der Malsburg C. 342.
 Von Grünau M. 150, 158.
 Von Neumann J. 263, 265-268, 270-
 272, 279, 281, 283, 295, 297, 315,
 318, 324, 330, 331, 335, 337, 374.
 Von Uexküll J. 553.
 Vosberg R. 26.
 Waller F. 295.
 Walton K. 106, 452, 531.
 Warhol A. 437.
 Warren R. 295, 427.
 Warrington E. 310.
 Wasserman G. 208, 209.
 Waterhouse L. 325.
 Weinberg D. 177.
 Weiner P. 13.
 Weinstein S. 13.
 Weiskrantz L. 406, 413, 414, 420,
 454.
 Welch R. 486.
 Werner J. 131.
 Wertheimer M. 149.
 West L. 26.
 White S. 13, 457.
 Whiten A. 283, 553.
 Wiener N. 223.
 Wiesel T. 58, 59.
 Wilkes, K. 546, 551.
 Wilson D. 12, 244, 283, 314.
 Winograd T. 122.
 Wittgenstein L. 87, 88, 399, 426, 483,
 509, 554, 572-574.

Wolfe J. 346.
Woodfield A. 13.
Wordsworth W. 35, 37.
Wundt W. 62.
Wyeth A. 476.
Wynes K. 13.

Yonas A. 225.
Young J. 282.
Zajonc R. 498.
Zeki S. 173.
Zihl J. 412.

Index des matières

- Absence (de saisie du *petit mal*) 439, 440.
 Absence de représentations vs représentation de l'absence 444.
 Absolutisme 522, 563.
Abstracta 127, 453.
 Abstraction 512, 513, 519, 521, 529.
 Abus, sexuel 521.
 Accélération 52.
 Accès 288.
 – immédiat 396.
 – limité 463.
 – moins qu'optimal 245.
 – (pas dans le processus de production) 332-334.
 – privilégié 87, 94, 160, 306-308.
 – de Shakey vs le nôtre 124, 386.
 Accumulateur 266, 295, 296.
 ACT* 331-339, 349.
 Acte de langage, voir aussi Discours 103-105, 157, 214, 286, 294, 295, 298, 299, 307, 308, 311, 312, 375, 390, 392, 451, 567.
 Acte gratuit 207.
 Action, intentionnelle 48, 312, 313.
 Activation, seuil d' 210.
 Adaptation à la vision inversée 486-493.
 ADN 40.
 Affabulation 311.
 Affect 64, 69, 82-88.
 Agent, rationnel 103.
 Aléatoire, roman 32.
 Algorithme 280.
 Allégorie 113, 124.
 Ambiguïté de portée 304, 308.
 Âme, 49, 533, 560-562.
 – immortalité de l' 453, 454.
 – théorie des centres de gravité 453.
 Amnésie 311, 489.
 Amusement 87.
 Amusement, comme variable 87.
 Amygdale 59.
 Analyse-par-synthèse 24.
 Analyseur cortical 441.
 Anesthésie
 – feinte 58.
 – locale 196.
 Anosognosie 105, 312, 439, 441, 442.
 Antérieur, cingulum 141.
 Anthropologie 110-115, 126, 127, 318-320, 323.
 Anticipation 224-226, 229, 236.
 Antimatière 54.
 Anton, syndrome d' 105, 443, 444.
 Aphasie 310, 311, 383.
 Apparence, et réalité 151-153, 169, 170.
 Apparence réelle 48, 50, 169.
 Appétit épistémique 29, 228, 406, 439-443, 473.
 Appréciation 48.
 Appréhension, innée 379.
 Apprentissage 242, 243.
 – comme réorganisation 231.
 – par imitation 274.
 – par instruction explicite 274.
 Araignée 514-516, 518, 547.
 Araignée, chemins pour 576.
 Architecture de tableau 432.
 Archive, dossier d' 432.
 Argent 39.
 Argot 305.

- Arithmétique, opération 266.
 Arrière plan de la conscience 418, 419, 438, 448, 505, 580.
 Artefact (plan d') et plan naturel 222.
 Articulation, niveau supplémentaire d' 379.
 Ascidie 224.
 ASCII, code 296.
 Association 279-281, 479-481.
 Attention 278, 279, 475.
 - capture 238.
 Auteur 451, 453, 518.
 - apparent 109.
 - autorité de 111, 128, 305, 379, 394.
 - d'actes de langage 105, 287-289.
 - d'enregistrement 287.
 - intention de l' 109, 112.
 Automate 49, 99, 321, 404, 406, 516.
 Automatiste 320.
 Avortement 563.
- Baldwin, effet 230-236, 239, 245, 249, 259, 260.
 Ballistique
 - actions 186.
 - événement non rapportable 419.
 - nature des saccades 446.
 « Bas en haut » (de) 25.
 Base-ball 308.
 Baud, taux de 424.
 Behaviorisme 58, 96, 231, 573.
 - grossier 97, 199.
 Bibliothèque, la vision comparée à une 445.
 Bière, comme goût acquis 490, 491.
 Biofeedback 413.
 Bit 272.
 - carte à 303, 318, 320, 364-366, 369, 370, 431, 432, 438.
 - défini 265.
 Boîte en carton (des Rosenberg) 466, 473, 496.
 Boîte noire 218, 286.
 Boîtologie 353.
 Bon et mauvais 220, 221, 224, 237, 259.
 Bonne Vieille Intelligence Artificielle 337.
 Boucle, infinie 334.
 Brainstorm (film) 483, 484.
 Branchement conditionnel 266, 332.
 Bridge 374.
 Broyage 63.
 Bureau ovale 138, 140, 516, 533.
- Buts 334.
 Bureaucratie d'homoncles 295, 296, 298, 300, 304, 312, 327, 328, 533.
 Butoir
 - mémoire 271, 274.
 - visuel 358.
- Cache-tampon 414-420.
 CAD système 361, 363-365, 367, 389, 390, 398, 429.
 CADAVEUGLE 363-365, 368-371, 389, 429-432, 463, 464, 500, 501, 545.
 Cadavre, traitement des 562, 563.
 Cadre 329.
 Cadre, tampon de 432.
 Cantate 479, 480, 547, 548.
 Capitale, peine 563.
 Carrousel 211-213, 578.
 Cartésien (voir Descartes, R.)
 - goulot 140.
 - interactionnisme 51.
 - matérialisme 141, 155, 179, 185, 320, 400, 401, 427.
 - Théâtre 31, 57, 141, 142, 146, 148-151, 158, 164, 166, 170, 172, 176-179, 183-185, 200, 207-210, 214, 215, 285, 287, 290, 318, 320-322, 348, 357, 358, 365, 371, 378-381, 388, 389, 392, 401, 402, 427, 440, 450, 451, 458, 486, 492-494, 523, 526, 535, 539, 546, 552, 565, 568, 571.
 Castor 238, 514, 515.
 Catégorie (erreur de) 68.
 Causalité
 - inobservable 171.
 - mécanique 199.
 Cause et effet, ordre de 193, 194.
 Cécité 65, 340, 421-424.
 - déni de, voir syndrome d'Anton 105, 408, 409.
 - hystérique 105, 407, 408.
- Centre
 - du cerveau (où l'expérience est supposée se produire) 57, 139, 140, 142, 146, 185, 186, 207-210, 296, 327.
 - de gravité 126-128, 135, 454, 512, 518, 519, 569.
 Centre de Gravité Narrative 507, 508, 518, 519, 529, 532, 533, 535, 544, 555, 556, 561, 562.
- Cerveau
 - pense-t-il ? 45, 46.
 - comme machine 48.
 - comme esprit 29, 50.

- dans une cuve 15-20, 31, 546.
- électronique géant 268, 271.
- perle (de cerveau), voir aussi Perle de l'esprit 453-455, 523-527, 534.
- Cerveau divisé, patients 248, 327, 524-529.
- Chambre Chinoise (expérience de pensée de la) 402, 540-546.
- Champ, magnétique 177.
- Chaotiques, processus 231.
- Chat (quel effet cela fait d'être un) 547.
- Chauve-souris 546-556.
- Chef
 - dans le cortex préfrontal 343.
 - dans le thalamus 342, 343.
- Chimpanzé 238, 243, 572.
- reconnaissance de soi par le 530.
- Choix forcé 408-410, 412, 579.
- Cinéma 22.
- Cinéma vérité* 22.
- Clares et distinctes (idées) 352.
- Code pour la couleur (voir aussi codage des couleurs) 433, 461, 499, 500.
- Coqito ergo sum* 15, 92.
- Colombier (de Platon) 275-279, 333, 348, 349, 353, 375.
- Comme si 452, 453.
- Commissurotomie 525, 528.
- Communication
 - dans le cerveau 247, 262, 342, 351, 365-367, 375, 388-390, 393, 400, 405, 406.
 - verbale, et son importance dans les expériences 99-101.
- Compétition dans le cerveau 298, 343, 421, 441, 545, 546, 557.
- (voir aussi évolution, réseau Le Gagnant Ralle Tout)
- Comportementaux, critères 503.
- Compréhension 77-80, 348, 544, 545.
 - du langage 95.
 - où se produit-elle ? 401, 539.
 - phénoménologie de la 77-80.
- Compression, algorithme de 432.
- Compte rendu 180, 287-290, 298, 315, 320, 379, 380, 395.
- Concentration 346.
- Conception (*design*) 353.
- Conception vs imagination 351, 352.
- Concepts, phénomènes qui dépendent de leurs 39.
- Conceptualiseur 293, 294, 296, 298, 300, 312, 313.
- Conflit, résolution de 333.
 - programme de 238, 239.
- Connaissance 353.
- Connexion, force (comme microhabitude) 272.
- Connexionnisme 336-338.
- Conscience, *passim*, voir aussi Flux de conscience
 - comme mode d'action, et non comme sous-système 210.
 - des animaux 31, 83, 98-100, 547, 556.
 - concept de 38.
- Construction (logique) 110.
- Contenu (voir aussi Intentionnalité) 166, 438, 451, 567, 568.
 - de l'expérience 93, 94, 183.
 - déterminé 296, 301, 392.
 - états individualisés par le 397.
 - fixation de 451, 535, 566.
 - indéterminé 307.
 - passant par les neurones 329.
 - qualitatif 460.
 - véhicules du 170.
- Contention scheduling 238, 239.
- Contrepèterie (spoonerism) 292.
- Contrôle (voir aussi Autocontrôle) 207, 214, 227, 236, 237.
 - flux de 296.
 - usurpé, par délégué 301.
- Contre-intuitif, caractère 54-56, 138, 281, 512, 534.
- Corps calleux 525.
- Cortical, analyseur 441.
- Couleur 428, 460-477, 564, 571.
 - aveuglement aux 469.
 - cellules au sujet de 340.
 - codage des 366, 465-467, 476, 477.
 - discrimination des, par le cerveau 172, 189.
 - expérience de la 44.
 - par les nombres 429, 430, 436, 437, 500, 501.
 - phi 149-151, 157-159, 164, 165, 171, 175, 180, 199, 450.
 - solide 433.
 - vision de 47, 76, 106, 226, 433-437, 493-495.
- CPU 272, 273, 295.
- Créativité 305, 372.
- CRT (tube à rayons cathodiques) 263, 269, 358, 363, 364, 367, 368, 371, 378, 580.
- Croyance 104, 385, 386.
 - environnement 560-562.
 - expression de 105, 114, 125, 130, 168, 169, 380, 381, 467.

- fixation 327.
- erreur, expression de 114.
- Curiosité 29, 261, 262.

- Dame Nature 221, 222, 225, 229-231, 327, 341, 349, 468, 513, 514.
- Déchiffrement de l'esprit, scientifique 44.
- Dégoût 513, 514.
- Délibération 40.
- De l'intérieur » 60.
- Démon 238, 239, 298-301, 313, 329-332, 333, 340, 344, 351, 352, 382, 443, 508, 544-546.
- (voir aussi Homoncule)
- Démystification 37, 40.
- Déplacement Doppler 136, 551, 552.
- Désir 104, 558.
- Dessin 73, 74.
- Détecteur d'yeux 446-448, 453, 454.
- Détection de collision 228.
- Deus ex machina 40.
- Développement comme réorganisation 231.
- Deviner quand deviner 412-414.
- Dichromatique, vision 549, 550.
- Diffusion radiophonique dans l'esprit 246, 322, 347.
- Discipline
 - des échecs 372, 373.
 - du langage 374.
- Discours
 - compressé 185.
 - perception 185.
 - production 291, 292, 297, 300, 392.
 - sons 69, 70, 101.
 (voir aussi Acte de langage)
- Discrimination par le cerveau 172, 173.
 - distribuée 371.
 - état discriminateur 195, 196.
 - discriminatoire (crudeness) 226.
 - qui n'a besoin que se produire une fois 165, 166, 195, 196.
- Disposition 461, 462, 478-481.
 - propriété dispositionnelle 464, 470, 571.
 - réactive 481, 484-487, 489-493, 500, 501.
- Distinction hémisphère gauche/droit 228, 247, 268, 526.
- Donné 171.
- Double prise 349.
- Douleur (voir aussi Souffrance) 40, 45, 82, 83, 86, 90-92, 113, 127, 128, 172, 200, 312, 395, 396, 478, 556-562, 565.
 - est-elle réelle ? 571.
 - projetée 167.
 - psychosomatique 408.
 - rapportée 382, 384.
 - rêvée 83, 84.
- Drogues (effet des) 26.
- Dualisme 43, 50-59, 139, 164, 172, 199, 200, 320, 427, 442, 443, 543, 544.

- Échéance 191-196.
- Échecs 492-495.
- Écholocalisation 547-554.
- Écriture, importance dans la formation de la conscience 274.
- Ectoplasme 53.
- Édition dans le cerveau 147, 157, 173, 184, 195, 196.
 - chambre d'édition 157-159, 194, 202.
 - changements éditoriaux vs errata 163, 307.
- Effets secondaires 222.
- Ego 31, 285.
- Egoïsme, variétés de 221.
- Egoïque 219.
- Élan vital 40.
- Electromyogramme 207.
- Empirisme (britannique) 76, 81, 92.
- Encapsulation 300, 327.
- Énergie 52.
- Entraînement
 - de l'oreille 419.
 - du cerveau 272.
 - du palais 491.
- Épilepsie 439, 440, 524-526.
- Épiphénoménisme 97, 493, 495, 572.
- Épiphyse 51.
- Épithélium 46, 47, 65.
- Epochè 62.
- Équateur 126.
- Esprit
 - comme nid à mèmes 259.
 - comme structure perçue par un esprit 383.
 - âme ou moi comme 525, 526.
 - esprit/cerveau 57.
 - est le cerveau 50.
 - œil de l'esprit, œil mental 73, 368, 371, 386, 433, 440, 460, 463, 481.
 - oreille de l'esprit 80, 83.

- perle mentale 453, 454.
- substance 50, 51, 53, 114, 427.
- Erreur 306, 307.
- Espace
 - couleur solide comme exemple d' 433.
 - du représentant (véhicule) vs du représenté 184, 185.
 - logique 168, 169.
 - personnel 420.
 - phénoménal 168, 379, 440.
 - problème 334, 335.
 - qualité 472, 473, 486, 491, 500, 501.
 - sémantique 307.
 - de travail global 321, 339.
 - visuo-moteur 491, 492.
- Espèce naturelle 509.
- Essentialisme 522.
- Esthétique, jugement 81.
- Éther 53.
- Et que se passe-t-il ensuite ? 319, 330-343.
- Éthologie, cognitive 553.
- Être à propos de (voir Intentionnalité)
- Évolution 37, 83, 84, 215-285, 296, 302, 346, 366, 476, 495, 496, 507, 520, 534.
 - culturelle 242.
 - dans le cerveau 231, 232, 241, 299.
 - de la vision des couleurs 467-473.
 - de la douleur 82.
 - des moi 512-514.
 - et souffrance 559.
 - vitesse de l' 235, 236.
- Existentialisme 207, 208.
- Expérimentale, méthode 407.
- Expériences, conception des 99-101, 104.
- Expérimentation animale 563, 564.
- Expliquer, par opposition à éliminer 562-564.
- Exploration (voir aussi Auto-exploration) 558.
- Explosion combinatoire 214, 301, 307, 381.
- Expression (voir aussi Croyance, expression de) 380.
- Faire vs arriver 49.
- Fantôme 21, 519.
 - dans la machine 50, 327, 512, 526.
- FAP (*fixed action pattern*) 325.
- Feedback 306, 307, 414, 417, 419.
 - clôre la boucle 412.
- Fenêtre
 - contrôle 214.
 - édition 202.
- Fetch-exécute cycle 267, 280, 331, 332.
- Fiction 451, 452, 506, 542.
 - et mondes fictionnels 105-109.
 - grammaticale 573.
 - interprétation de 123.
 - littéraire vs du théoricien 126-130.
 - traîtée comme un fait 113.
- Fictionnel, caractère du moi 508, 532.
- Figment 428, 433, 434, 437, 439, 444, 451, 452, 458, 459, 463, 539, 565.
- Figure et fond 417.
- Fitness* 249, 253, 256.
 - dépendante de la fréquence 258, 259.
 - des mêmes 275.
- Fixation d'organisation postnatale 231, 249.
- Flux de la conscience 64, 92, 148, 174, 175, 178, 189, 210, 238, 268, 279, 295, 318, 322, 323, 440, 504, 505, 540.
- Fonction 220.
 - de la conscience 275, 344, 345.
 - de la vision 406.
 - multiple 222, 223, 340, 341, 345.
- Fonctionnel, vs épiphénoménal 495-501.
- Fonctionnalisme 47, 406, 477, 569-571.
 - des machines de Turing 570.
 - homonculaire 329.
 - téléofonctionnalisme 570.
- Font attention » neurones qui 198-203.
- Formation réticulée 141, 342.
- Formulateur 294-297, 300.
- Fovéa 75, 440.
- Frontière 512, 513, 516, 517, 560-562.
 - biologique 221, 223.
 - de luminosité 95, 580.
 - entre moi et le monde extérieur 142, 221, 223.
- Futur, producteur de 185, 223-225, 237, 347.
- Gadget 349.
- Gagnant Raffle Tout, réseau 238, 239.
- Galvanique, réponse de la peau 96.

- Gène, comme entité théorique 97.
 Génotype 229.
 Gorille 238, 248.
 Grain (niveau de) 124.
 Grammaire, comme source de discipline 374, 375.
 Grammaticale, structure 70.
 Gravité (voir aussi Centre de Gravité Narrative) 229.
 - centre de 126, 127, 135, 454, 512, 515, 571.
 - effet de, sur la lumière 136.
 Gremlins 499, 500.
 Guillemets 63.
 Guitare 68.
- Habitude 548.
 - de l'ordinateur, déterminée par le logiciel 269, 270.
 - mauvaise, de pensée 317, 393, 568.
 - microhabitude 270.
Hack, odd (jeu de mots avec *ad hoc*) 278.
 Hallucination 15-32, 59, 154, 155, 298.
 - auditive 315.
 - de mémoire 155.
 Harmonique d'un ton au-dessus 67-69.
 Haute fidélité, enregistrement et playback 67-69.
 Hémianopie 403.
 Hémnégligence 440-442.
 Hétérophénoménologique, monde 114, 124-127, 130, 157, 160, 162, 199, 286, 287, 307, 367, 372, 397, 408, 452, 481, 504, 505, 508, 555-557, 571.
 Hétérophénoménologie 98-106, 112, 261, 262, 288, 289, 330, 407, 445, 451, 491, 518, 539, 560, 568.
 - des animaux 551-562.
 - neutralité de 114, 126-130.
 Heuristique 349.
 Hilarité 84, 85.
 Histoire, raconter des 323, 375, 518, 531.
 Holisme 338.
 Hominien 238-240, 243, 244, 246, 268.
Homo ex machina 116.
Homo sapiens 240, 243, 252, 260, 327, 342, 515.
- Homogénéité des propriétés 459, 460.
 Homoncule 27, 28, 116, 117, 121, 125, 223, 282, 311, 326-329, 405, 439, 441, 532, 565.
 Horizon de simultanéité 140.
 Hypnogogique, effet 374.
 Hypnotique, suggestion 84.
 Hypothèses, test des 25, 26.
 Hystérique, cécité 407, 408.
- IBM PC 272, 273.
 Idéalisation 568, 569.
 Idée
 - de rouge 459-463.
 - selon Descartes et Hume 60.
 - simple 251.
 Identité, théorie de l' 198, 201, 569-271.
 Illusion
 - bénigne 451.
 - de la conscience 536, 538, 540, 565.
 - de l'utilisateur 269, 273, 387-390.
 - des couleurs 464, 465.
 Illusionniste (voir aussi Magique, Truc) 22, 23, 26.
 Image 77, 78.
 - de Shakey 116-126, 383-387.
 - mentale 114, 124, 125, 168, 169, 206, 207, 287, 319, 372, 432.
 - réalité des images mentales 569-571.
 - réelle 122, 123.
 - rotation de 357-371.
 Imagination 43, 44, 68, 83, 452, 492, 534, 556.
 - blocage 480.
 - comment elle est construite 76.
 - crutch for 363, 400, 545.
 - échec de 30, 71, 72, 350-352, 378, 458, 493-495, 534-539, 543-545, 558.
 - extension de 29, 114, 318, 351, 352.
 - limites de 547.
 - rôle dans l'amusement et la souffrance 554.
 Immanence, illusion 445.
 Immortalité 533, 534.
 Immunitaire, système 221.
 Immunologie 241.
 Impasse 334.
 Impressionnisme 62, 75, 76.
 Inconscient 296.
 - action (initiée inconsciemment) 203, 206, 207.

- buts 303.
- cognition 350, 351.
- conduite auto 176.
- contrôle 408, 409.
- ou politiques automatisés
- pensées d'ordre supérieur 383-385, 387.
- pensée 382-384.
- processus
- production du langage, compréhension 129.
- slander 113.
- Incorrigibilité 90, 92, 397.
- Indétermination des mondes fictionnels 109.
- Infailibilité
 - de l'introspection 92.
 - papale 112, 127.
- Information
 - combien il y en a peu dans la vision aveugle 407-409, 411, 424-443.
 - mesure de (voir aussi largeur de bande, taux de baud, bit) 18.
 - notre capacité d'utiliser l' 29.
 - passant de l'extérieur à l'intérieur 75, 76.
 - traitement de l', chez les chauve-souris 551, 552.
 - transmission et traitement dans le cerveau 41-43, 185, 187, 537-538.
- Informationnelle, complexité 465, 495.
- Informativore 228.
- Informel, neutralité de 225.
- Ingénierie
 - génétique 99.
- Initiative de Défense Stratégique 192.
- Inné
 - biais dans l'espace des qualités 486.
 - connaissance 333.
 - mécanismes du langage 240, 249, 273, 374.
 - mécanismes et structures 225, 229, 273, 319, 326, 329, 480.
 - tendance 49.
- Instruction, registre de 267, 280.
- Instrumentalisme 569-571.
- Intention
 - communicative 243, 293, 298-304, 306, 309, 311, 395-397.
 - consciente 205-208.
 - de l'auteur 113, 305.
- Intentionnalité 103, 104, 241, 414-417, 567, 568.
 - définie 414.
 - intrinsèque 348.
- Intentionnel
 - action 47, 211.
 - objet 110, 126, 130, 169.
 - posture 103, 104, 568.
- Interactionnisme 51, 59.
- Interprétation 105, 129.
 - auto- 306.
 - de la fiction 113.
 - de textes 106-108, 428, 429, 451, 452.
 - des états du cerveau par lui-même 398, 399.
 - du discours 101-105, 286.
 - par le cerveau 146, 209.
- Intrinsèque, propriété 85-87, 460, 461, 474, 478, 480, 491, 506, 533, 539, 557, 558, 567.
- Introspection 30, 75, 76, 90, 91, 125, 276, 289, 319, 383, 384.
 - comme impliquant des pensées du troisième ordre 381, 382.
 - comme théorisation 92, 93, 125.
 - de la machine à la von Neumann 268, 269.
 - du moi 511, 512.
 - limitation de 437, 438.
- Introspectif, piège 445.
- Introspectionnisme en psychologie 63, 82, 96.
- Intuition 413, 458, 487.
 - pompe à 353, 493-496, 542, 548.
- Intelligence artificielle (IA) 57, 80, 123, 232, 238, 239, 269, 271, 275, 280, 302, 319, 320, 323, 328, 329, 331, 336, 349, 544, 560.
- Interne, œil 393.
- Interne, lumière, la conscience comme 555.
- Inverser les lunettes 486, 492, 493.
- Invisibilité de la structure fonctionnelle 261, 262, 273.
- IRM (*innate releasing mechanism*) 325.
- Isotropie 327, 328, 349.
- Jargon, comme exprimant de nouveaux concepts 263.
- Jargon, aphasie du 310, 311.
- « Je » 529.
- Jeu de mot 303.
- Joycienne, machine 267, 272, 279, 280, 287, 343-350, 555.
- Jugement 166, 397, 402, 427, 450-452, 460.

- de simultanéité et de succession 210-212.
- Juger et paraître 172, 173.

- Kinesthésie 17, 65, 66.
- Kludge 263, 278.
- Korsakoff, syndrome de 311.

- Label 365, 436-439.
- Laffer, courbe de 143.
- Lamarckisme 235, 259.
- Langage 281, 461.
 - assembleur 376.
 - compréhension du 95.
 - créativité du 372.
 - de la pensée 376-378.
 - effet sur la structure cérébrale et la compétence 249, 258-262, 376.
 - évolution du 239, 240, 244, 249.
 - machine 273, 376.
 - mécanismes innés 239, 240, 249, 273, 374.
 - naturel comme langage de programmation 376.
 - perception vs production 290.
 - production (voir aussi production du discours) 379.
 - programmation 295.
 - et protolangage 243.
 - rôle dans la conscience 30, 278, 374, 375, 553, 554.
 - rôle dans la création du moi 515-518.
 - rôle dans la machine virtuelle 274.
 - par signes 303.
- Lapin, cutané 183, 184, 199.
- Lapsus 303.
- Latéralité, test de 177, 178.
- Lecteur d'épreuves, effet du 427, 428.
- Leibniz, loi de 570.
- Liage, problème du 156, 322, 323.
- Libre arbitre 39, 40, 207, 211, 533.
- Ligne cachée, retrait de 400.
- Lieux, liés (loci) 258, 259.
- Linguistique 290-292, 310.
- Lisp 269, 296, 377.
- Lobe frontal 342, 343.
- Locus ceruleus 458.
- Logiciel (software) 272-274, 565.
 - défini 263.
 - dans le cerveau 240, 262.
 - niveau 539, 545.
- Logique, de la croyance et de la connaissance 567, 568.

- Lorentz, équations de 177.
- Lumière, vitesse de la 136, 140.
- Luminescence 494, 504, 505.

- Machine
 - langage 295, 377, 378.
 - personne comme 534-536.
 - table 266-268.
- Madame Bovary* 107.
- Magie, scène (voir aussi Tour de passe-passe) 23, 164, 348, 349, 446, 539.
- Main-œil, coordination 237.
- Mal, pourquoi les douleurs font 83.
- Maladie 557-562.
- Marie, la spécialiste des couleurs 493-508.
- Matérialisme 40, 50, 53, 54, 60, 140, 180, 188, 189, 204, 402, 493, 500-505, 533, 543.
- Matérialisme éliminatif 508.
- Maxwell, équations de 177.
- Mélodie, discrimination de 578.
- Membre fantôme 17, 21.
- Même 249-262, 275, 303, 319, 328, 330, 375, 475, 478, 517.
- Mémoire 297, 300.
 - à accès aléatoire 265.
 - à court terme 203.
 - à long terme 203.
 - butoir 271.
 - comme bibliothèque 158, 171.
 - couleur 463.
 - défectueuse 395, 396.
 - de l'ordinateur pas comme du cerveau 273, 278.
 - de travail 331-333.
 - épisodique 347.
 - hallucination 204.
 - perte de 311.
- Mémosphère 257, 258, 273.
- Mental, équipement 297, 298, 301.
- Mentalais (voir aussi Langage de la pensée) 290, 294, 451.
- Mentalistes (termes) 56, 57.
- Mentalité d'avant l'écriture 274.
- Métaconnaissance 425, 543.
- Métacontraste 181-183.
- Métaphore 113, 121-124, 168, 228, 289, 290, 343, 358, 387-390, 565.
- Micro-habitude 270, 272, 319.
- Milliseconde 137.
- Miracle 56, 299, 319, 565.
- Modularité 327.
- Moi 30, 45, 206, 207, 273, 282, 348, 507, 508, 512-534.

- biologique 512-518, 530, 555.
- comme âme 335.
- comme non indépendant des mêmes qu'il arbore 259, 377.
- connaissance 556, 557.
- contrôle du 275, 345, 346, 517, 518, 529-533, 555.
- exploration de 261, 262, 282, 319, 374.
- gestion de 275, 311, 398, 411.
- localisation du 207-209.
- observation du 392.
- reconnaissance du 530-533.
- représentation du 387, 518, 533.
- sondage du 214.
- stimulation du 245-248, 260, 272, 278, 343, 375-377, 558.
- stimuli en file (cuing) 413.
- transparence du 375.
- Moindre effort, principe de 314.
- Monde (voir aussi Monde hétérophé-
noménologique)
- fictionnel 168.
- hétérophénoménologique 109.
- Montrer, vs dire 369-371, 393.
- Moral, agent 39, 40.
- Morale, responsabilité 290, 291.
- Moralité, relation à la conscience 556-
565.
- Morse, code 70.
- Mots, connaissance des 387, 388.
- Mouvement apparent 172, 420.
- MPD (Multiple Personality Disor-
der) 520, 521, 523, 524, 526, 527.
- Multiple, fonction 340, 341, 345.
- Mystère 54, 55, 59, 383, 384.
- défini 36.
- de la conscience 28, 30, 31, 37,
38, 40.
- faire face au 350.
- systématique 503.
- Mystère, nouveaux adorateurs de 341.
- Mythe 562-564.
- Narcissisme 473.
- Narration, voir Récit
- Navigation 135, 191-193.
- Necker, cube de 360.
- Négligence 439-444.
- bénigne 441, 448.
- financière 442, 443.
- typographique 441.
- Néo-Laffer, courbe 145, 489.
- Neuro-anatomie 342.
- Neurochirurgie 196, 197, 211.
- Neurochirurgien fou 484, 485, 489.
- Neuronale, adéquation 198, 200, 202,
203.
- Neurophilosophie 202.
- Neurosciences 29, 44, 57, 97, 188,
320-322, 329, 337, 465, 473.
- hypothèse de base des 96.
- Neutralité 99, 114.
- de l'hétérophénoménologie 98,
112, 130.
- Neutrino 53.
- Niveau d'explication 344-346.
- NMDA récepteur 342.
- Nobel, prix 59, 320.
- Noumène 63.
- Nourriture, recherche de 238, 239.
- Nouvelle-Orléans, bataille de la 188-
190, 214.
- Observateur 135, 136, 168, 175, 290,
317, 320, 502.
- dans le cerveau 140, 210, 388,
389, 391, 534.
- localisation de l' 141, 163.
- rôle du 470-472.
- Observation
- connaissance sans 398, 399.
- interne 397.
- Observation des oiseaux 417, 418.
- Occipital, cortex 403-405.
- Occurrence, propriété 339, 340, 478,
480.
- Olfaction 65.
- Ontologie 53.
- Opérationnalisme 127, 128, 164.
- première personne 169-171.
- Opinion, opposée à la croyance 131.
- Optimalité, hypothèse de 354.
- Orchestre, petit dans l'esprit 66.
- Ordinateur 187, 188.
- asynchrone 187, 188.
- parlant 103, 105, 123-125, 187,
188, 263-273, 287, 540-546.
- Ordre-supérieur, états intentionnels
383, 390.
- Ordre supérieur, pensée d' 392, 393,
395, 413, 414.
- Organisation (*design*) voir Conception,
Plan.
- Orientation, réponse d' 227, 237-239,
276.
- Orwellien vs Stalilien 151-162, 164-
166, 171, 183, 203-205, 207, 213,
397, 488, 546, 573.
- Ouïe 66-71.

- Où-dire, comme modèle d'accès 393.
 Où tout se rejoint 57, 141, 172, 173, 209, 371, 453.
- Pandémonium 237-239, 275, 297, 298, 300-302, 312, 318, 328, 379, 392, 418, 565, 578-580.
- Paradoxe 30.
 - de la conscience 135.
- Parallèle
 - architecture 261, 262, 334.
 - dans le cerveau 146, 148-151, 172, 173, 261, 262, 268, 269.
 - simulé 271, 332, 333.
 - traitement 238, 270, 297, 298, 303, 312, 318, 326, 337.
 - traitement distribué 335-337.
- Paranormal 409, 519.
- Parasite 513.
 - même comme 256, 275, 319.
- Parenthèses (mise entre) 63.
- Parler à soi-même 243-246, 275, 277, 372, 504.
 - à quoi cela sert ? 375.
 - par opp. à des systèmes se parlant à eux-mêmes 393.
- Pathologie 308-312, 342, 343, 402, 440-444.
- Paysage, adaptatif 232-236.
- PC-Paintbrush 431.
- Pensée 30, 394.
 - comme discours à soi-même 82.
 - d'ordre supérieur 391, 392, 395, 413, 414.
 - inconsciente 382.
 - opposée la croyance 382, 383.
 - penser en pensées 372.
 - sans image 82.
- Pensée, expérience de 16, 29, 30, 67, 403, 453, 484, 485, 492, 493, 494, 540, 541, 546.
 - comme pompe à intuition (voir Intuition)
- Penseur le, de Rodin 278.
- Perception de la profondeur 140.
- Percu, vs inféré (mouvement) 160.
- Perpétuel, Machine à Mouvement 52.
- Phénomène 110-115.
- Phénoménal
 - champ 30, 71, 93.
 - propriété 402, 420, 422, 423, 460, 474.
 - espace 168, 379, 440.
- Phénoménologie 63, 64, 66, 75, 83, 86-92, 113, 127, 172, 200, 312, 320, 369, 433, 538.
 - comme comportement d'une boîte noire 215.
 - de la compréhension 77.
 - du rire 85.
 - paradoxale 160.
 - pure et impure 95, 96.
 - « réelle » 451, 452.
 - visuelle 74-76.
- Phénoménologie (comme école philosophique) 62, 348.
- Phénéthio-urée 470.
- Phénotype 229.
 - étendue 514.
 - évolution dans le 231-233.
- Phi, phénomène 149, 576.
- Philosophes 15, 30, 37, 59, 86, 98, 99, 115, 149, 153, 319, 320, 348, 351, 352, 403, 406, 457, 458, 481, 494, 495, 502, 522, 538, 540, 541, 546, 567-547.
- Phonème, restauration de 428.
- Physicalisme (voir aussi Matérialisme) 495.
- Physique, système à symboles 353.
- Physique 53, 54, 199.
 - révolution 460.
- Piano, accordage 419.
- Pinéale, glande 138, 141.
- Pixel 340, 370, 371, 432, 437.
- Plan (*design*) voir Conception, Organisation.
- Plan incliné 194, 201.
- Plannification 225, 425.
 - opportuniste 303.
- Plasticité
 - des ordinateurs 263.
 - des systèmes nerveux 229.
 - phénotypique 233-236, 242, 260, 273.
- Platon, colombier de (voir Colombier)
- Platoniciens, mêmes 256-258.
- Point aveugle 403, 404, 427, 428, 432, 439-441.
- Point de vue 135, 136, 163.
 - changement de 423.
 - créé par la réplique 221, 223.
 - de l'observateur 223.
 - smeared 141, 175, 194.
- Point de vue du plan 353.
- Politique, choix 416, 419.
- Pôle auditif du barbier 95.
- Poltergeist 52.
- Pontifical, neurone 282, 512.
- Positivisme 482.
- Postal, tampon 188, 194, 201.
- Posture du plan (*design stance*) 353.

- Posture intentionnelle 103, 104.
 Potentiel évoqué primaire 200, 201, 205.
 Poules, préférences des 558, 559.
 Pouvoir, et non fonction, des mécanismes de la conscience 346.
 Pré-expérientiel vs post-expérientiel 142, 154-156, 160, 307, 487-489.
 Pré-cognition
 - écartée 150.
 - par carrousel 211, 212, 578.
 Préconscient 314.
 Préfrontal, cortex 141, 342, 343.
 Première personne 91, 92.
 - opérationnalisme 171.
 - perspective (ou point de vue) 91, 92, 96, 160, 163, 418.
 Présentation 141, 172, 173, 195, 196, 214, 215, 319, 320, 388-390, 450, 486, 552.
 - (voir aussi Re-présentation).
 Pressentiment 425, 450-452, 567.
 Préverbal, message 294, 295, 298, 300, 306.
 Primitive, opération d'ordinateur 264-267, 295, 331.
 Prise de décision 207, 301, 540.
 Prise en compte 48, 59, 220, 556-565.
 Problème épistémologique des autres esprits 547, 553.
 Problème, primordial 224.
 Production, système de 331-334.
 Programme
 - comme description d'une liste d'instructions 269.
 - comme liste d'instructions 331.
 - comme recette 270.
 - de la machine virtuelle de l'esprit 272.
 Projection 165.
 - à rebours dans le temps 166, 167, 169, 179, 200.
 - dans l'espace 167-169.
 Projecteur, théorie de l'attention comme 342.
 Prologue 372.
 Prompting, rôle dans la vision aveugle 408-410.
 Proposition 450, 451.
 Propriété
 - dispositionnelle 459-462, 464, 470.
 - occurrente 459-461, 478.
 Prosopagnosie 95.
 Prothésique, vision 420-424.
 Psychanalyse (jeu de société) 23-29, 298.
 Psychologie 57, 99-101, 319, 320, 369, 495.
 Psychologie populaire 378, 382, 385, 389-398.
 Psychophysique 473, 474.
 Psychosomatique (cécité) 408.
 Publication (comme mauvaise métaphore pour la conscience) 163, 164.
 Qualia 30, 427, 442, 443, 454-509, 538, 539, 557, 565, 570, 571.
 - inversés 87, 419.
 Qualité, première et seconde 459, 462, 465, 469-473.
 Qualités jolies et suspectes 469-471.
 Qualophile 478, 480, 482, 485-487.
 Quatre fonctions (les) 237.
Quel effet cela fait-il ? (what is it like) 125, 127, 130, 238, 239, 479, 480, 527, 528, 547-556.
 R et D 260.
 Radar 530-532.
 Raisonnement, pratique 312, 313.
 Raisons, naissance des 220, 221, 223, 513.
 RAM 265.
 Rapporter 287-290, 298, 313, 322, 394.
 - vs exprimer 195, 378-384, 392, 395.
 Rapportabilité 418, 419.
 Rationalité 313.
 - hypothèse de 568.
 - idéale 375.
 - minimale 375.
 Récit 517, 518.
 - cadre 545.
 - création de 305.
 - de soi 306, 530.
 - flux 148.
 - explication évolutive comme 216.
 - flashback dans un 189.
 - fragment de 173, 319, 326.
 - observateur comme 176.
 - ordre de composition 189.
 - précipité par des sondages 173-175, 184, 214.
 - rêve 26, 27.
 - séquence de 175.
 Reaganéconomie 143.

- Réalisme 569-572.
 Réalité, des objets intentionnels 112.
 Récursion 320, 384-386.
 Réductionnisme 85, 479, 564, 565, 569-571.
 Réentrée, cartes de, circuits 336.
 Référence à rebours dans le temps 198, 200, 201.
 Réflexe 49, 136, 159.
 Réflexion 400, 411, 556.
 Registre 272.
 - mémoire 267.
 - nombres, utilisés pour représenter 433.
 Régression à l'infini
 - comme l'arrêter 395.
 - des Signifieurs 299, 300.
 Relativisme 472.
 Rendu 364-366, 370, 439, 450, 463.
 Renforcement 274.
 Répétition 278, 345-347, 531.
 Replicateur 220-222, 251-254, 256-258.
 Représentation 240-242.
 - dans le cortex 339.
 - minimale 225.
 - par le système immunitaire 221.
 Représentation en gros continu 432, 437.
 Représentants, Chambre des 340, 341, 441.
 Représenté vs représentant (véhicule) 108, 109, 169, 176, 184, 185, 188-190, 193, 204, 206-208, 210.
 Res cogitans 45, 46, 50, 59, 140.
 Rêve 23, 24, 27, 28, 84, 299.
 Rêve éveillé 82, 83.
 Rire 85-87.
 Robot 534, 556.
 - animal 62, 560, 561.
 - conscient 536, 540.
 - réplicateur 222.

 Saccade 74, 146, 228, 416, 438, 439, 446-448, 579, 580.
 - suppression saccadique 446-448.
 Savants 54.
 Scène, voir Truc, Tour de passe-passe.
 Schizophrénie 315.
 Science cognitive 30, 57, 199, 295, 320-322, 327, 495.
 Scotome 404-406, 411, 413, 439.
 Script 323, 419.
 Secondaires, qualités (voir Qualités)
 Seconde loi de la thermodynamique 256.

 Seconde intention 542.
 Secret 553-555.
 Sémantique
 - analyse 101-103.
 - préparation 173.
 - similarités de niveau 108.
 Sémiotique, matérialisme 508, 509.
 Sensa 460, 461.
 Sensation 30, 73.
 Sensation brute 402, 460.
 Sense data 402.
 Sentiment 420.
 Séquence subjective 175, 197, 210, 214, 215, 577.
 Sériel
 - chaînage 275.
 - processus 265, 268, 272, 295, 299, 301, 313, 319, 323-325, 330.
 - recherche 347.
 Serpent, haine des - chez les primates 477.
 Sexe 225
 - évolution du 218, 219.
 SI-ALORS, primitive (voir aussi Branchement conditionnel) 331, 332.
 Signification 102-104, 293, 294, 299, 304, 305.
 Signifieur 304, 306, 308.
 Signifieur Central 282, 290, 293-295, 299, 306, 307, 310, 314, 322, 378, 450-452, 523, 535, 538, 565.
 Silicone, machine en, par opp. à organique 47.
 Simulateur de vol 398.
 Simultanéité, subjective 141, 175, 206, 209.
 Singe
 - préférence de couleur du 477.
 - vervet 243.
 Singularité 177.
 Slogans, aller à la chasse aux 571.
 Soar 331, 334-337, 339, 349.
 Soi, voir Moi
 Solipsisme 499.
 Somatosensoriel, cortex 197-199.
 Somnambulisme 48, 49.
 Son, vitesse du 136, 140.
 Son, studio 69, 70.
 Son, bande 147, 194.
 Sondage 148, 173, 184, 214.
 - pas de temps optimal du 175.
 Souffrance 557-563.
 Sourd et muet 377, 556.
 Souvenir 246, 347, 369, 425.
 Spécialisés, systèmes, dans le cerveau 227, 237, 244, 245, 276, 286,

- 300, 318, 319, 322-326, 330, 342, 346, 432, 433, 445.
- comme généralistes 339-441.
- Spectre inversé, voir Qualia inversés.
- Stalinien, voir Orwellien vs stalinien
 - modèle de Libet comme 200, 203, 207.
- Star Wars* 536.
- Stéréo, son 168.
- Séréogramme à points aléatoires 168.
- Subjectif, contour 434.
- Subjectivisme 472.
- Subjectivité 171, 202, 460, 481-483.
- Substantia nigra 458.
- Sujet
 - hétérophénoménologique 166, 169.
 - unifié 104, 105.
- Surface de l'esprit conscient 277.
- Surface de réflectance spectrale 464.
- Suspectes, qualités 469, 470.
- Symbole, amovible 353.
- Sympathie 84.
- Système expert 47.
- Systèmes, réponse des (voir Chambre Chinoise).
- Systèmes cérébraux dorsaux et ventraux 545.
- Tableau noir (architecture du) 331.
- Tachistoscope 181.
- Tampon dateur 188, 194.
- Taste-vin 47.
- Télétransport 534.
- Télévision 18.
 - vitesse de 137.
- Témoignage 393.
- Témoin 402, 442, 523, 535, 565.
 - substance mentale comme 44, 45.
- Temps
 - au bout du doigt 205.
 - comment il est représenté dans le cerveau 185-195.
 - fenêtre 155, 185, 193, 194.
 - pression du 185, 192, 193.
 - réel 165, 190.
- Termite 514, 515.
- Thalamus 59, 342.
- Théâtre, organe de l'esprit 68.
- Timing, mesure du
 - absolue 204, 205, 207-210, 212.
 - des événements cérébraux 148, 195, 196, 209.
 - des événements conscients 148, 162.
- quand elle détermine le contenu 192.
- Toucher 66.
- Tour de passe-passe, voir aussi Magie 166, 351, 352, 387, 414.
- Tout simplement comme ça, histoire 243.
- Trace, garder la 366.
- Traitement central 327, 328.
- Traitement de texte 388, 389.
 - le cerveau comme 333.
 - l'ordinateur non conçu comme 264.
- Traitement parallèle distribué (voir aussi Connexionnisme) 336, 337.
- Trame 568.
- Transcription 101-103.
- Transformation de Fourier 67.
- Tremblement de terre 39.
- Trichromatique, vision des couleurs 467.
- Troisième personne, perspective à la 96-98, 127, 166, 418, 548, 552.
- Trou noir 53, 217.
 - apprendre un nouveau 238, 239.
 - Bon Truc 232-235, 240, 246, 248-250, 26, 283, 468.
 - comme entité théorique 97.
- Truc
 - apprendre un nouveau 238, 239.
 - Bon Truc 232-235, 240, 246, 248-250, 260, 283, 468.
 - conjurer (voir aussi Magie, Scène, Tour de passe-passe) 164, 573.
 - de conception naturelle 49.
 - hallucination comme moment où le cerveau fait un tour de passe-passe 21.
 - heuristique comme sac à 349.
 - jouer un tour de passe-passe avec le temps 198.
- Végétarien 564.
- Vérificationnisme 164, 170, 482, 499, 500, 571, 572.
- Verticale, détection de symétrie 226, 237.
- Veto émis par le moi 207, 213.
- Vidéo, jeu 18.
- Vidéo, bande 432.
- Virtuel
 - câble 245.
 - capitaine 282.
 - espace 360, 364.
 - langage de programmation 377.

- machine 261, 263, 269-274, 278-280, 282, 319, 324, 326, 337, 350, 387, 388, 535, 539, 540, 543, 555, 565.
- objet 357.
- structure 375.
- Virtus dormitiva* 86.
- Viser 414-416.
- Vision 71-76, 347, 348, 402-440.
 - par ordinateur 116-123.
 - localisée à l'intérieur des yeux 142.
 - parafovéale ou périphérique 74, 94, 438, 445, 447, 448, 579, 580.
 - résolution de la 65.
- Vision aveugle 309, 402-414, 419, 421, 425, 442, 443, 546, 579, 580.
 - apprentissage sous 412.
 - chez les chauve-souris 551.
- Vision dichromatique 553.
- Visualisation 367-369, 452.
- Vitalisme 40, 351, 352.
- Vitesse de communication 271.
- Voir 415-418, 420, 439.
 - en tant que compréhension 77.
- Voyance 165.
- West Side Story* 108.
- Wordperfect 273.
- Wordstar 269, 273.
- Zimbo 386-388.
- Zombie 98, 99, 105, 112, 126, 350-353, 376, 383-387, 390, 502, 506, 558, 572.
- Zombiste 320.
- Zoologie 62-64, 91, 96.

Table des matières

Préface	11
Prélude.....	15

Première partie Problèmes et méthodes

<i>Chapitre 1 : Expliquer la conscience</i>	35
La boîte de Pandore : la conscience doit-elle être démystifiée ?	35
Le mystère de la conscience.....	41
Les attrait de la substance mentale.....	43
Pourquoi le dualisme est caduc.....	50
Le défi.....	57
 <i>Chapitre 2 : Visite au jardin phénoménologique</i>	61
Bienvenue au phénôme.....	61
Notre expérience du monde extérieur.....	64
Notre expérience du monde interne.....	76
Affect.....	82
 <i>Chapitre 3 : Une méthode pour la phénoménologie</i>	91
La première personne du pluriel	91
La perspective à la troisième personne.....	95
La méthode de l'hétérophénoménologie.....	98

Les mondes fictionnels et les mondes hétérophénoménologiques.....	105
Le charme discret de l'anthropologue.....	110
À la découverte de ce dont il est vraiment question.....	112
Les images mentales de Shakey.....	115
La neutralité de l'hétérophénoménologie.....	126

Deuxième partie

Une théorie empirique de l'esprit

<i>Chapitre 4 : Les Versions Multiples contre le Théâtre</i>	
Cartésien.....	135
Le point de vue de l'observateur.....	135
Introduction du modèle des Versions Multiples	146
Les révisions orwelliennes et staliniennes.....	151
Le théâtre de la conscience revisité.....	164
Le modèle des Versions Multiples en action	172
<i>Chapitre 5 : Le temps et l'expérience.....</i>	179
Moments flottants et lapins bondissants.....	180
Comment le cerveau représente le temps.....	185
Le « renvoi en arrière dans le temps » de Libet	195
Les arguments de Libet sur les retards subjectifs d'intentions.....	206
Un bonus : le carrousel précognitif de Grey Walter.....	211
Derniers détails à régler.....	213
<i>Chapitre 6 : L'évolution de la conscience.....</i>	217
À l'intérieur de la boîte noire de la conscience	217
Les débuts.....	219
Scène une : la naissance des frontières et des raisons...	219
Scène deux : comment produire le futur mieux et à nouveaux frais.....	223
L'évolution dans les cerveaux et l'effet Baldwin	230
La plasticité dans le cerveau humain : le décor	236
L'invention de bonnes et de mauvaises habitudes d'autostimulation	243
Le troisième processus évolutif : les mêmes et l'évolution culturelle.....	249

Les mêmes de la conscience : la machine virtuelle à installer	260
<i>Chapitre 7 : Ce que les dires peuvent nous faire.....</i>	285
Faisons le point : <i>e pluribus unum</i> ?.....	285
Bureaucratie contre Pandémonium.....	290
Quand les mots veulent se faire dire.....	302
<i>Chapitre 8 : L'architecture de l'esprit humain</i>	317
Où en sommes-nous ?.....	317
Comment s'orienter avec le croquis schématique.....	321
Et que se passe-t-il ensuite ?.....	330
Les pouvoirs de la machine joycienne.....	343
Est-ce bien une théorie de la conscience ?.....	350

Troisième partie

Les problèmes philosophiques de la conscience

<i>Chapitre 9 : Montrer et dire.....</i>	357
La rotation d'images dans l'œil de l'esprit.....	357
Mots, images et pensées.....	371
Rapporter et exprimer	378
Les zombies, les zimbos et l'illusion de l'utilisateur	385
Les problèmes de la psychologie populaire.....	391
<i>Chapitre 10 : Le programme de protection du témoin démantelé</i>	401
Résumé	401
La vision aveugle : le fait d'un zombie partiel ?	402
Cache-tampon : un exercice de prise de conscience.....	414
La vision prothésique : à part l'information, qu'est-ce qui manque ?	420
« Remplir » contre trouver.....	426
La négligence comme perte pathologique d'appétit épistémique.....	440
La présence virtuelle.....	444
Voir c'est croire : un dialogue avec Otto.....	448

<i>Chapitre 11 : Les qualia disqualifiés.....</i>	457
Un nouveau fil de cerf-volant.....	457
Pourquoi y a-t-il des couleurs ?	464
Apprécier nos expériences.....	475
Une fantaisie philosophique : les qualia inversés.....	482
Des qualia « épiphénoménaux » ?	493
Pour en revenir à mon rocking-chair.....	503
 <i>Chapitre 12 : La réalité des moi.....</i>	511
Comment les êtres humains tissent un moi.....	512
Combien de moi par client ?.....	519
L'insoutenable légèreté de l'être.....	529
 <i>Chapitre 13 : La conscience imaginée</i>	535
Imaginer un robot conscient	535
Quel effet cela fait d'être une chauve-souris ?.....	546
Matière à penser	556
La conscience expliquée ou éliminée ?.....	563
 <i>Appendice A (pour philosophes)</i>	567
<i>Appendice B (pour scientifiques).....</i>	575
<i>Bibliographie.....</i>	581
<i>Index des noms</i>	605
<i>Index des matières</i>	611

La Conscience expliquée

Qu'est-ce qui fait d'un peu de matière un être animé ? Qu'est-ce qui confère à certaines structures physiques le privilège énigmatique d'éprouver des sensations et d'avoir des expériences ? La conscience.

Mais que sait-on aujourd'hui de la conscience ? Est-elle une donnée immédiate comme le veulent la tradition et le bon sens ? Ou bien repose-t-elle sur des processus en grande partie inconscients que des machines pourraient reproduire ?

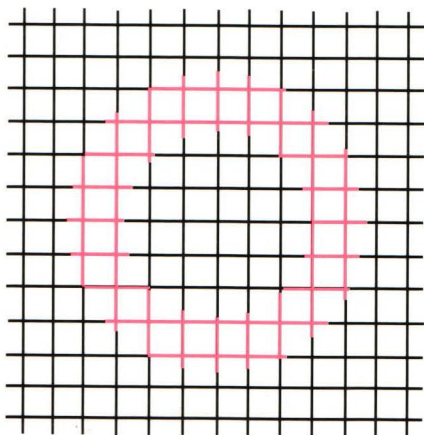
Daniel C. Dennett se livre ici à une brillante réfutation des théories traditionnelles de la conscience, et propose un nouveau modèle fondé sur les acquis modernes de la psychologie, de la neurologie, de l'informatique, et de l'intelligence artificielle.

Daniel C. Dennett

Daniel C. Dennett dirige le Center for cognitive Studies de Tufts University. Il est l'un des chefs de file de la philosophie de l'esprit aux États-Unis.

Traduit de l'anglais (États-Unis) par Pascal Engel

*Existe-t-il un emplacement dans votre
cerveau où le rose que vous voyez ici est
« rempli » de façon à produire en vous
une expérience consciente ?
(voir pages 426-440)*



ISBN 2.7381.0216.6



37,35 € 245 F

www.odilejacob.fr