

LES SCIENCES COGNITIVES APPLIQUEES

Daniel Andler, Elena Pasquinelli

TABLE DES MATIERES

| | |
|--|-----------|
| 1. Appliquer les sciences cognitives : qu'est-ce à dire ? | 1 |
| 1.1. Pertinence des sciences cognitives selon le domaine d'activité..... | 1 |
| 1.2. Niveau théorique des interventions : un continuum de profondeur..... | 2 |
| 1.3. Aller-retour et recherche translationnelle..... | 3 |
| 1.4. La relation spéciale entre les sciences cognitives et le numérique | 3 |
| 1.5. La dimension éthique | 4 |
| 2. L'individu en interaction avec le monde | 4 |
| 2.1 Ergonomie, design, interfaces | 4 |
| a. Le rôle croissant des sciences cognitives dans la conception des artefacts | 4 |
| b. Ergonomie cognitive..... | 5 |
| c. Interface homme-machine (IHM) | 5 |
| d. Design interactif | 6 |
| e. Les sciences cognitives appliquées au design et à l'architecture d'espace | 7 |
| f. Analytique visuelle | 7 |
| 2.2. Le monde augmenté | 8 |
| a. Réalité virtuelle, interfaces multi-modales..... | 8 |
| b. Objets, surfaces, agents intelligents, réactifs, connectés : la réalité augmentée | 9 |
| 3. Les agents cognitifs, la cognition externalisée | 9 |
| 3.1 L'intelligence artificielle | 9 |
| a. Quel est le rapport entre IA et sciences cognitives ? | 9 |
| b. L'IA de l'autonomie à la complémentarité | 10 |
| c. En pratique, comment l'IA se situe-t-elle aujourd'hui ? | 11 |
| 3.2 L'apprentissage automatique | 12 |
| a. L'apprentissage en général et la place de l'apprentissage automatique | 12 |
| b. Types et systèmes d'apprentissage automatique | 13 |
| c. Apprentissage automatique, intelligence artificielle et sciences cognitives | 13 |
| 3.3 La robotique cognitive | 14 |
| 3.4. Le traitement automatique du langage naturel | 15 |
| 3.5 La vision artificielle | 16 |
| 4. La cognition réparée ou augmentée par le biais de son organe | 17 |
| 4.1 Les interfaces cerveau-machine | 18 |
| 4.2 Les neuroprothèses sensorielles..... | 18 |
| 4.3 De la thérapie électro-convulsive à la stimulation cérébrale profonde..... | 19 |
| 4.4. L'entraînement cognitif | 19 |
| 4.5 La stimulation cérébrale à la maison et les nootropiques chimiques..... | 20 |
| 5. Contribution des sciences cognitives aux processus sociaux | 20 |
| 5.1 Applications ponctuelles..... | 21 |
| a. Les limites de l'attention et la conduite automobile..... | 21 |
| b. La fiabilité de la mémoire et les témoignages oculaires | 21 |
| c. Les nudges ou comment orienter les choix des gens sans leur faire violence..... | 22 |
| 5.2 Justice, commerce, politique, éducation | 22 |
| a. Le tournant cognitif de la justice | 23 |
| b. Le tournant cognitif du commerce..... | 23 |
| c. Le tournant cognitif de la politique | 24 |
| d. Le tournant cognitif de l'éducation | 24 |
| 6. Les sciences cognitives dans le contexte social | 26 |
| 6.1 Distorsions et exagérations : les neurosciences cognitives source de fantasmes | 26 |
| a. Des méthodes pseudo-scientifiques..... | 26 |
| b. Les neuromythes | 27 |
| c. Quels antidotes ? | 27 |
| 6.2 Questions d'éthique | 28 |
| a. Le principe de l'adaptation à l'utilisateur | 28 |
| b. Apprentissage automatique et données massives..... | 29 |
| c. Implants cochléaires et culture sourde | 29 |
| d. Les nudges et le paternalisme libertarien | 30 |
| e. L'augmentation cognitive..... | 30 |
| f. La neuroéthique | 31 |

Les sciences cognitives ont désormais atteint une certaine maturité, qui se traduit par leur capacité à se déployer dans un grand nombre de domaines d'activité : elles sortent volontiers des laboratoires pour se pencher sur toutes sortes de processus, proposer des réponses aux questions des spécialistes des domaines concernés, en poser de nouvelles, contribuer à la mise au point de nouvelles façons de faire, de nouveaux instruments, appareils, dispositifs matériels, produits, et de bonnes ou meilleures manières d'utiliser ces objets innovants.

La plupart des secteurs d'activité sont potentiellement concernés, pour la raison que l'agent humain joue partout un rôle central et que ce rôle repose sur ses capacités cognitives. En pratique cependant, les sciences cognitives n'interviennent aujourd'hui de manière visible que dans certains domaines. Leur apport consiste, selon les cas, en une prise en compte du fonctionnement cognitif réel, permettant de concevoir de meilleures méthodes, par exemple en matière de choix, d'éducation ou de pratiques médicales, juridiques, économiques... ; en une analyse des interactions entre l'agent humain et son environnement, tout particulièrement les artefacts, signes, machines, appareils qui l'entourent, analyse permettant d'améliorer ces interactions ; en la mise au point de prothèses permettant de pallier des déficits sensoriels ou moteurs ; en la conception de systèmes de traitement de l'information et de communication adaptés au fonctionnement individuel et collectif des agents et capables d'y apporter des améliorations.

Ce chapitre présente un échantillon représentatif de ces interventions, ainsi que des questions pratiques et éthiques qu'elles soulèvent. Certaines interventions sont spécifiques au domaine d'activité considéré : la navigation (le repérage de l'espace et la détermination du chemin à suivre) concerne au premier chef l'architecture, les transports, l'urbanisme... ; les processus d'apprentissage à différents âges concernent le secteur éducatif ; l'adhésion thérapeutique concerne la médecine ; la fidélité du témoignage visuel concerne la justice, etc. Mais beaucoup d'interventions sont transversales, même si elles peuvent donner lieu à des déclinaisons spécifiques à différents secteurs ; de manière générale, ce sont celles qui concernent l'information et la communication, que celles-ci passent ou non par le canal d'appareils et de systèmes artificiels. Il n'y a donc pas une seule manière de classer les champs ou modes d'application des sciences cognitives. L'ordre que nous suivrons dans ce chapitre reflètera surtout le type d'intervention, mais il se référera aussi parfois au domaine d'application.

Ce chapitre est long, mais il est composé de segments qui peuvent se lire indépendamment les uns des autres. Le §1 pose sur un plan théorique général la question de l'application en tant que telle, et certains lecteurs pourront choisir d'en remettre la lecture à la fin. Les § 2 à 5 portent sur différents types d'application, et sont relativement indépendants entre eux. Le §6 aborde les dimensions culturelles et éthiques des sciences cognitives.

1. Appliquer les sciences cognitives : qu'est-ce à dire ?

1.1. Pertinence des sciences cognitives selon le domaine d'activité

Toute activité implique l'agent humain, seul ou (plus généralement) en groupe, ce qui fait, on l'a dit, que le développement de toute activité dépend en principe de nos connaissances sur le comportement humain, autrement dit pour une large part des sciences cognitives. Or on n'a pas attendu les sciences cognitives pour commencer à agir : de fait, toutes les activités que nous connaissons, à l'exception significative de ce qui touche aux TIC (technologies de l'information et de la communication), nous y reviendrons, sont largement antérieures à l'émergence des sciences cognitives. De là à penser qu'en dépit des apparences, elles ne servent en réalité à rien, il n'y a qu'un pas. Certains le franchissent, le présent chapitre devrait leur prouver qu'ils ont tort.

Inutiles ou indispensables ? La réalité se situe entre les deux, et occupe en fait un spectre continu. Examinons d'abord rapidement les deux pôles. À première vue, nous n'avons pas eu besoin de psychologie du raisonnement ou de neurosciences de la décision pour découvrir le théorème de Pythagore, la composition de l'eau ou le bacille de Koch, pour choisir de passer nos vacances en Bretagne ou de préparer le concours de Polytechnique ; pas davantage que pour enseigner les mathématiques ou gagner le tournoi de Wimbledon ; pour inventer le régulateur de Watt ou instituer le suffrage universel. À l'autre extrême, lorsque nous cherchons à réaliser une rétine artificielle ou un implant cérébral pour la commande motrice directe par la pensée chez un quadriplégique, ou encore un poste de pilotage d'avion de chasse, nous n'imaginerions pas de ne pas consulter les travaux pertinents de psychophysique et de neuroscience.

Ce qui permet d'agir sans sciences cognitives dans le premier cas est une combinaison de deux facteurs : d'une part nous possédons une théorie « naïve », une compréhension non scientifique des mécanismes mentaux à l'œuvre (nous n'en avons évidemment aucune des mécanismes neuraux, mais ce sont leurs effets fonctionnels qui nous semblent seuls pertinents, simplification admissible dans la plupart des cas mais pas

tous) ; d'autre part nous bénéficions d'un considérable apprentissage par l'expérience, aux échelles de l'évolution, de la culture et de l'individu. (Cet apprentissage peut nous conférer à la fois, directement, la compétence requise pour telle ou telle activité, et la maîtrise de notre théorie naïve des processus mentaux, laquelle à son tour nous donne ou renforce cette compétence.)

Théorie naïve et compétence pratique acquise par l'expérience ont cependant des limites. Ces limites apparaissent dès lors que les conditions dans lesquelles se situe l'activité considérée s'écartent de celles qui prévalent, ou prévalaient normalement, ou que l'on quitte l'espace des cas habituels. Il arrive aussi qu'on s'interroge rétrospectivement sur ce qui se passe en réalité également dans les cas habituels et dans les circonstances normales : les explications que nous avons considérées comme exactes, ou allant de soi, sont peut-être insuffisantes, voire carrément fausses ; ou bien nos procédures standard pourraient être améliorées à la lumière de connaissances plus exactes des mécanismes impliqués. Ainsi, l'apprentissage de la lecture par un enfant dyslexique peut être inopérant selon les méthodes habituelles, et les sciences cognitives interviennent pour suggérer une bonne méthode ; mais on peut du coup se demander si les méthodes habituelles, appliquées aux enfants neurotypiques, sont optimales, et si on ne pourrait pas les améliorer, réduire l'effort et le taux d'échecs, par exemple, en mobilisant ce que les sciences cognitives nous ont appris.

Entre les cas où nous avons, peut-être à tort, l'impression de comprendre ce que nous faisons, et de le faire bien, sans le secours des sciences cognitives, et les cas où nous attendons tout d'elles, il existe un continuum de cas dans lesquels nous avons l'impression de comprendre un peu, mais pas assez, pas assez pour faire aussi bien qu'il le faudrait, et pour résoudre les problèmes plus difficiles qui se posent dans notre pratique. Pour prendre un exemple, nous avons des intuitions sur le fait que souvent les gens favorisent le court terme, alors qu'un raisonnement simple leur montrerait qu'il serait de leur intérêt de viser le long terme ; ou sur le fait qu'ils généralisent à tort à partir d'exemples qui leur sont familiers ; ou qu'ils cherchent à tort des explications aux coïncidences, sous-estimant systématiquement la possibilité d'occurrences peu probables ; ou qu'ils prennent une corrélation pour une cause. Ces erreurs de jugement affectent profondément la vie économique, sociale et politique, et nous voudrions les éliminer ou du moins en limiter les effets nocifs, par exemple en modifiant la présentation des données, l'organisation des choix, ou en proposant des logiciels d'aide à la décision. Mais pour y parvenir, nous avons besoin d'une compréhension précise que seules les sciences cognitives peuvent nous apporter.

1.2. Niveau théorique des interventions : un continuum de profondeur

D'un domaine à l'autre, et souvent même au sein d'un même domaine, la contribution des sciences cognitives varie considérablement. On a affaire ici aussi à un spectre continu, cette fois en matière de profondeur théorique, ou de richesse de contenu.

Au pôle minimal, les sciences cognitives ne font guère plus que d'attirer l'attention sur la place de l'agent « réel » dans les processus étudiés ou dans l'utilisation d'artefacts, et sur le fait que cet agent « réel » n'est pas nécessairement identique à l'agent théorique que l'on suppose dans le domaine considéré. Dans le même ordre d'idées, les sciences cognitives, sans apporter d'informations spécifiques, peuvent suggérer par leur simple exemple d'appliquer des méthodes empiriques sanctionnées par la science, d'adopter une méthodologie plus rigoureuse, de procéder à la collecte plus systématique de faits pertinents, de déjouer certains pièges classiques dans le raisonnement, notamment en matière de statistique et probabilité et de causalité. Ce mode d'intervention concerne en premier lieu la recherche théorique dans d'autres disciplines, telle que l'économie, et ne rentre pas dans le périmètre du présent chapitre. Mais des domaines tels que l'éducation, la santé ou la politique, qui ont souvent systématiquement négligé, pour différentes raisons, l'apport potentiel des méthodes empiriques, commencent à en comprendre l'utilité : les approches « fondées sur les faits » (*evidence-based*) y ont désormais droit de cité.

Un cran au-dessus se trouvent des situations dans lesquelles l'apport des sciences cognitives consiste en concepts. Les exemples les plus simples sont ceux d'information, de représentation, de computation ; certains sont plus spécifiques : mémoire de travail, charge cognitive, spécificité de domaine, valeur par défaut... Mais il faut franchir un pas supplémentaire pour voir apparaître des apports empiriques des sciences cognitives, des connaissances précises sur certains mécanismes mentaux ou cérébraux, ou sur certaines contraintes formelles ou matérielles sur ce qu'on peut attendre d'un dispositif naturel ou artificiel de traitement de l'information. Les sciences cognitives fournissent alors non pas une théorie développée, elles ne formulent pas des recommandations impératives, mais elles font office de boussole, de source d'inspiration, et leur représentant sur le terrain joue le rôle d'un consultant.

Le pôle maximal est atteint lorsqu'un phénomène ou un artefact du domaine d'application est complètement théorisé par les sciences cognitives : ce qui est produit est la contrepartie concrète d'une théorie, dont les spécialistes des sciences cognitives et du domaine d'application évaluent conjointement les

résultats, un peu à la manière dont physiciens et ingénieurs mettent au point ensemble un instrument de mesure, par exemple un scanner. Ces cas ne se réalisent qu'au prix d'une démarche collaborative d'un genre particulier, comme nous allons le voir.

1.3. Aller-retour et recherche translationnelle

Comme c'est toujours le cas lorsqu'une discipline académique se déploie dans le monde, l'application des sciences cognitives n'atteint son plein régime que si elle résulte d'un processus bidirectionnel : elles offrent des ressources aux secteurs d'activité qui font appel à elles, mais ceux-ci à leur tour leur posent des questions auxquelles elles n'avaient pas pensé, dans des termes qui ne leur sont pas familiers ; en sorte que dans un deuxième temps, il est nécessaire de s'engager dans une recherche dite translationnelle, qui consiste précisément à identifier les questions intéressantes pour le secteur en question et à définir les méthodes permettant aux sciences cognitives d'y répondre. Ce type de recherche a été mis à l'honneur dans le domaine médical : comment passer de manière productive de la paillasse du laboratoire de biologie au chevet du malade (« *from bench to bedside* »), et inversement ?

Pour les sciences cognitives, le problème est d'assurer une circulation féconde entre les connaissances qu'elles ont acquises sur les mécanismes à l'œuvre dans la cognition, d'un côté, et les processus propres aux différents secteurs qui peuvent en bénéficier et qui peuvent en retour leur apporter de nouvelles questions, de nouveaux concepts et des données empiriques. Ce travail est engagé, généralement depuis peu, et se fait naturellement secteur par secteur, car il implique une connaissance intime du domaine d'application d'un côté, de la partie potentiellement concernée des sciences cognitives de l'autre. C'est un dialogue de spécialistes, qui seul peut conduire à des interventions « profondes », au sens de la section 1.b.

1.4. La relation spéciale entre les sciences cognitives et le numérique

Si les sciences cognitives sont comprises comme l'étude des systèmes biologiques de traitement de l'information, et les technologies numériques comme la conception de systèmes artificiels de traitement de l'information, on est fortement tenté de penser qu'on a affaire aux deux faces d'une même médaille, ou plus précisément à un schéma en V dont la pointe est occupée par une théorie abstraite du traitement de l'information et les deux autres sommets par l'étude des systèmes biologiques et la conception des systèmes artificiels respectivement.

Ce schéma a en effet été proposé lors de l'émergence conjointe des sciences cognitives et de l'intelligence artificielle au cours des années 1950 (v. le chapitre « Les sciences cognitives : d'où viennent-elles... »). Mais même dans ce cadre, il n'était crédible que moyennant une restriction importante, à savoir que les systèmes artificiels considérés remplissent des fonctions spécifiques (telles que raisonnement, mémoire, apprentissage, coordination motrice, planification, reconnaissance de formes...) et une fonction générale (intelligence), analogues, sinon rigoureusement identiques, à celles de la cognition humaine (ou, si l'on souhaite élargir aux systèmes cognitifs d'espèces non humaines, à celle de la cognition animale). Un système dont la fonction serait, mettons, de produire un 0 quelle que soit l'entrée n'aurait guère de rapport avec un système biologique ; son concepteur n'aurait rien à apprendre des chercheurs en sciences cognitives, ni eux de lui. Mais d'un autre côté, il aurait été également improductif d'imposer une isomorphie parfaite entre le système cognitif humain et les systèmes artificiels qu'on cherche à construire. Il s'en est ensuivi un long débat, au cours des années 1960-1970, sur la bonne « fourchette » de similarité entre les deux types de systèmes.

Aujourd'hui, le schéma en V est, sinon abandonné, du moins ramené à la périphérie du domaine, pour au moins deux grandes raisons. D'une part, les progrès des neurosciences d'un côté, de l'informatique de l'autre, ont beaucoup réduit la portée des comparaisons entre code ou calcul neural et computation au sens logico-informatique. D'autre part, l'objectif de la simulation a largement cédé la place à celui de la coopération : selon l'opinion majoritaire des 30 dernières années, ce qui nous est le plus utile, et qui est davantage à notre portée, ce sont des systèmes artificiels qui *complètent* notre équipement informationnel naturel, plutôt que des systèmes artificiels *mimant* l'esprit/cerveau humain. (Nous assistons cependant depuis quelques années à un certain renversement de tendance.)

Il n'en demeure pas moins qu'il existe une affinité profonde, voire fondamentale, entre la cognition et ce que nous désignerons ici comme « le numérique », c'est-à-dire l'ensemble des processus rendus possibles par les technologies de l'information et de la communication (TIC). Due au fait que la cognition naturelle est depuis l'origine des sciences cognitives vue comme une affaire de traitement de l'information, cette affinité explique que les sciences cognitives interviennent à de multiples niveaux dans la sphère numérique.

1.5. La dimension éthique

Toute application d'un domaine scientifique dans un secteur d'activité humaine engage des enjeux éthiques. Les sciences cognitives ne font pas exception. Elles ne peuvent éviter de s'interroger sur l'usage de connaissances ou d'équipements qu'elles ont directement ou indirectement engendrés ; elles doivent fixer des priorités, elles doivent tenter de peser sur les politiques publiques, etc. Elles sont à cet égard dans la même situation que les autres grands champs disciplinaires. Mais de même qu'elles ont un rapport privilégié avec le monde *numérique*, en tant que science du traitement biologique de l'*information*, elles ont un rapport privilégié avec le monde *humain* en tant que science du traitement *biologique* de l'information. En choisissant de parler de traitement *biologique* (et non *humain* qui conviendrait également), nous entendons attirer l'attention sur deux dimensions essentielles : physiologique et neurale, évolutionnaire et comparative, ainsi que sur les deux niveaux de l'individu et de la société. Mais nous faisons également apparaître la proximité des sciences cognitives avec les sciences de la vie en matière d'éthique : c'est pour des raisons similaires que ces deux domaines ont une portée ou un engagement éthiques plus marqués que les sciences physiques ou les mathématiques, d'un côté, que les sciences humaines de l'autre. Pour les premières, du fait que celles-ci n'ont de rapport avec l'humain qu'indirect ou générique ; pour les secondes, du fait qu'elles n'ont pas la force persuasive des sciences de la nature : ce qu'un neuroscientifique ou un psychologue du développement peut dire, par exemple, de la dyslexie ou de l'altruisme, a de fait sinon de droit une autorité et donc une portée pratique très supérieures à ce que pourraient proposer la sociologie ou l'économie.

2. L'individu en interaction avec le monde

L'être humain interagit avec le monde par différents canaux : canaux directs — la perception dans ses différentes modalités et intermodalités, la motricité, canaux indirects — l'acquisition de connaissances, le langage... Le monde avec lequel il interagit est composite : il comprend des objets naturels, le plus souvent aménagés ou réaménagés par l'homme ; des êtres humains ; des artefacts, qui vont du repère signalétique aux objets manufacturés, de la cafetière à la perceuse électrique, à la voiture, à la pompe à essence ; parmi ces artefacts, des supports cognitifs tels que les livres et les enregistrements ; des systèmes informatiques ou informatisés, de la banque de données au smartphone, du robot domestique ou industriel au cockpit d'avion de chasse. Toutes ces choses interagissent entre elles, et l'individu dont elles constituent l'environnement interagit, séparément ou conjointement, avec chacune. L'individu ne s'engage pas du tout gratuitement dans ces interactions, mais parce qu'il est animé par des motivations générales, par les émotions du moment et par des intentions spécifiques : sa cognition « chaude » est impliquée au même titre que les composantes de sa cognition « froide », ce dont il faut tenir compte, autant que possible, dans les applications. Chaque sorte d'interaction donne lieu, au moins potentiellement, à des applications des sciences cognitives, dont nous ne pourrions présenter qu'un échantillon. Dans la présente section, nous parlerons d'applications qui visent essentiellement à améliorer les interactions avec l'environnement physique et les objets ; dans la suivante, de celles qui visent à compléter, enrichir, voire remplacer certaines fonctions cognitives en matière de connaissance et de communication, et d'action.

2.1 Ergonomie, design, interfaces

a. Le rôle croissant des sciences cognitives dans la conception des artefacts

La manière dont nous utilisons les outils, machines, appareils, instruments, meubles, ustensiles, armes... — bref, tout ce que nous fabriquons — est depuis les lointaines origines de l'humanité l'objet d'une intense activité : par l'observation, par la réflexion imaginative et par l'expérimentation, nous parvenons, au fil des générations, à améliorer, dans l'ensemble, ces artefacts : les résultats sont meilleurs, l'effort moindre, les accidents plus rares. Il en va de même de l'aménagement des espaces dans lesquels nous nous déplaçons, à toutes les échelles, de l'intérieur domestique aux bâtiments privés et publics, à la ville entière, au territoire...

La révolution industrielle, le développement des sciences de l'ingénieur, puis la fordisation, ont donné à ces efforts une impulsion décisive, en fournissant à la fois la motivation pour rechercher des solutions optimales (puisque une fois trouvé un dispositif optimal — dans l'absolu ou relativement à l'existant — on peut le reproduire à l'infini et à peu de frais, contrairement aux produits de l'artisanat, par exemple) et les moyens d'en trouver régulièrement, à savoir la méthodologie des sciences appliquées. L'optimal, il est important de le remarquer, prend de plus en plus en compte, au cours du XX^e siècle, la dimension esthétique : le *design* devient une discipline d'importance théorique et pratique égale à l'ingénierie industrielle classique, et va jusqu'à s'appuyer sur une doctrine selon laquelle la perfection fonctionnelle d'un artefact est constitutive de sa beauté.

Ce que les sciences cognitives apportent à cet effort multiséculaire, c'est tout d'abord une compréhension scientifique, beaucoup plus précise, plus détaillée, moins entachée d'erreurs, et surtout plus complète des mécanismes perceptifs et moteurs, des capacités de raisonnement, de mémoire, de conceptualisation, d'anticipation, etc. qui interviennent lorsque nous utilisons un artefact ou lorsque nous évoluons dans un espace aménagé : nous n'avons, sans leur apport, que des idées au mieux approximatives, souvent fausses et souvent aussi pas d'idées du tout.

C'est, ensuite, une nouvelle conceptualisation du *design*, appuyées sur une théorie des systèmes artificiels, due pour une bonne part à Herbert Simon, qui est à la fois l'un des fondateurs de l'intelligence artificielle (IA) et un pionnier de la psychologie cognitive. L'esthétique cognitive, c'est-à-dire la recherche des bases cognitives de la perception esthétique, constitue une autre source théorique importante.

Mais c'est surtout dans la conception d'outils ou de dispositifs qui incorporent des processus informationnels, c'est-à-dire cognitifs ou (si l'on préfère) proto-cognitifs, que les sciences cognitives interviennent de la manière la plus visible : si elles ne font que prendre le relais des méthodes traditionnelles en ce qui concerne les artefacts (simplement) matériels, s'agissant des artefacts numériques elles sont à elles seules la théorie de la pratique : tout ce qui se fait, concrètement, sur le terrain, en procède directement. Et comme de plus en plus d'artefacts incorporent des dispositifs numériques, la pertinence des sciences cognitives s'étend toujours plus loin (pour ne prendre qu'un exemple, pensons à celui de l'automobile). C'est cependant là seulement un idéal : en pratique, une bonne partie du travail se fait selon des procédures empirique par essais et erreurs, avatar numérique au XXI^e siècle des bonnes vieilles méthodes. L'intervention des « cognitiens », comme on les appelle souvent dans ce cadre, susceptible de changer la donne, ne prend de l'ampleur que progressivement. Mais elle est d'autant plus essentielle que les interactions impliquent des dimensions multiples : l'émotion et la motivation jouent un rôle qui ne le cède en rien aux capacités mnésiques, attentionnelles ou inférentielles. De fait, l'intrication des composantes « chaudes » à la théorie et à la pratique du design est un problème complexe dont la solution exige des ressources que les sciences cognitives sont les seules à développer.

Il existe donc un immense domaine, celui de la conception d'artefacts et d'environnements (y compris au sens spatial concret du terme) basée sur la considération de leur usage par des agents dont les capacités cognitives jouent un rôle fondamental. Ce domaine est structuré de manière assez floue : différentes spécialités, aux intitulés variables, se dégagent au fil du temps en fonction de leur insertion dans les circuits industriels et commerciaux et de la dynamique de l'innovation, plutôt que de leurs hypothèses théoriques, et souvent se chevauchent. On peut néanmoins distinguer des pôles caractérisés par le type d'objet ou d'environnement qu'il s'agit de concevoir ou d'agencer, nous allons maintenant en examiner quelques-uns.

b. Ergonomie cognitive

L'ergonomie est l'étude et la conception de postes de travail qui respectent autant que possible l'intégrité physique et mentale du travailleur, et qui favorisent sa productivité à dépense égale. Elle s'est longtemps focalisée sur les dimensions psychophysiques des tâches, en particulier dans les usines et les ateliers : éclairage, posture, gestuelle, acuité visuelle, auditive, tactile, fatigue, agencement des objets manipulés, examinés, touchés, vérifiés..., sur le plan de la pénibilité, de l'efficacité (rendement, taux d'erreurs, temps de remise en route après un incident...), de la sécurité du travailleur et de ses coéquipiers.

L'ergonomie est devenue « cognitive », à partir des années 1970, lorsqu'elle a enrichi ses modèles de l'agent : il n'est pas qu'un paquet d'os et de muscles, muni d'yeux et d'oreilles ; il est un système cognitif complet, capable de raisonner, de se souvenir et d'oublier, d'anticiper et d'évaluer, de planifier et de contrôler ; il a une conception des dispositifs sur lesquels il agit, et de son action sur eux ; il est capable d'apprentissage et d'adaptation, dans une certaine mesure et dans certaines conditions ; les routines réduisent son effort, mais trop de routine engendre l'ennui, etc. De tout cela il faut, du moins en principe, pouvoir tenir compte lors de la conception d'un poste, et plus largement d'un environnement de travail, ou quand on enquête sur des dysfonctionnements ou d'une souffrance des agents.

La plupart des postes de travail aujourd'hui sont plus ou moins complètement informatisés, et pas seulement dans le secteur tertiaire, puisque l'informatique embarquée est omniprésente dans l'industrie, les transports, l'agriculture, la santé, la sécurité et la défense. C'est donc sans surprise que l'ergonomie cognitive débouche sur un domaine beaucoup plus visible, celui des interactions avec les dispositifs informatiques.

c. Interface homme-machine (IHM)

Ce qu'en anglais on désigne par le sigle HCI, pour *human-computer interaction* est généralement traduit en français par « interface homme-machine » (IHM), alors qu'en anglais on distingue le HCI d'un champ un peu différent, le HMI c'est-à-dire justement *human-machine interaction*, appelé aussi il n'y a pas si longtemps MMI

pour *man-machine interaction*. Une complication supplémentaire vient du fait qu'on substitue souvent dans ces diverses locutions « interface » à « interaction » (que ce soit en français ou en anglais). Retenons de tout cela que l'objet de ce champ, conçu de la manière la plus large, est constitué par toutes les formes d'interaction entre des agents humains et des machines, sachant que lesdites machines comprennent non seulement des ordinateurs mais aussi des dispositifs ou des systèmes incorporant, de manière plus ou moins importante, des mécanismes informatiques de toutes sortes, les robots constituant un cas à part, et donnant lieu à une discipline spéciale, la robotique. Retenons aussi que les recherches portent à la fois sur l'observation et la compréhension des interactions et sur la conception d'interfaces entre l'agent humain et le système, interfaces qui sont eux-mêmes désignés plus précisément comme « interfaces homme-machine », ou « human-computer interfaces », d'où le glissement de termes (l'activité consistant à concevoir des interfaces X-Y étant parfois elle-même appelée « interfaces X-Y », ce qui n'a aucune importance).

Ce champ est présenté comme interdisciplinaire, relevant de quatre ou cinq disciplines principales et d'une kyrielle de sous-disciplines, du moins si l'on en croit l'article de Wikipedia. En réalité, ces disciplines et sous-disciplines sont elles-mêmes liées, voire confondues, et l'on doit préférer une caractérisation plus simple des « ingrédients » dont se compose le champ, que nous désignerons de son sigle français le plus courant, IHM. Ces ingrédients proviennent de quatre secteurs, qui sont (1) les sciences cognitives, conçues de manière large et incluant les approches comportementales, (2) l'informatique, (3) le design, (4) la sociologie, en y incluant tout ce qui touche aux média, à internet et aux réseaux sociaux, ainsi que les dimensions politiques et économiques. Ces secteurs sont de plus en plus profondément intriqués, surtout pour ce qui touche à l'IHM, et inversement l'IHM ne fait appel, bien entendu qu'à certains de leurs sous-domaines. Ajoutons aussi que, s'agissant d'interfaces particulières, soit du côté du système artificiel, soit du côté de l'agent humain, l'apport de disciplines spécifiques au domaine peut être nécessaire. De manière évidente, les interfaces dans une centrale nucléaire exigent des connaissances de physique et d'ingénierie, les interfaces dans les transports aériens des connaissances d'aéronautique et de droit aérien, les interfaces pour des malvoyants des connaissances en ophtalmologie, des interfaces pour la médecine de ville des connaissances médicales et de droit de la santé, et ainsi de suite.

Mais au cœur de toute activité de recherche et de conception en IHM est une réflexion sur les différentes interactions possibles entre un agent humain (générique ou présentant certaines caractéristiques spéciales) et un système artificiel, compte tenu à la fois des mécanismes chez l'humain et des réactions du système aux stimulations possibles.

C'est ainsi que le concepteur (designer) d'interfaces doit anticiper les différentes actions que l'utilisateur voudra mener et produire une interface qui répond aux besoins (usabilité), qui réduit les chances d'erreur (fiabilité), qui fonctionne même lorsque les conditions d'usage s'éloignent de celles qui sont initialement envisagées (robustesse), et qui donne lieu à une expérience positive (acceptabilité), ce qui implique la dimension affective (Peter & Beale 2008). Les connaissances issues des sciences cognitives peuvent guider le design – par exemple, en anticipant des problèmes d'utilisation liés aux limites de l'attention ou de la mémoire à court terme. En interaction avec designers et informaticiens, elles fournissent des modèles pour prédire les réactions de l'utilisateur face à certains aspects du design. Les sciences cognitives fournissent aussi des outils pour évaluer l'impact du design sur l'expérience, conduisant généralement à des recommandations concrètes, mais produisant aussi, graduellement, une attitude plus scientifique, critique et circonspecte, favorisant des stratégies à long terme.

Si le domaine de l'IHM se prête aux applications de la recherche en sciences cognitives, il alimente aussi, à l'inverse, la recherche fondamentale en nouvelles questions et en données. Un bon exemple de ce retour de l'appliqué vers le fondamental est la recherche sur le raisonnement visuo-spatial et le rôle des représentations graphiques dans l'accès à l'information, à la connaissance, et à l'apprentissage (Tversky 2011). En cherchant à mettre au point le meilleur design pour présenter une information visuelle, ce travail a identifié certains avantages cognitifs des représentations visuelles (en termes d'externalisation de certains processus de raisonnement, d'appui sur les capacités de perception visuo-spatiale et d'analyse spatialisée de l'information), mais aussi certaines limites (notamment le risque d'un excès de complexité et d'autres formes d'opacité).

d. Design interactif

Dans le droit fil de l'IHM, et en constituant de fait une branche, le design interactif (en anglais, *interactive design* ou *ID*) répond aux besoins croissants de penser les objets, généralement mais pas toujours entièrement informatisés, à partir de l'expérience de l'utilisateur : au lieu de partir de l'artefact, conçu pour remplir une certaine fonction, et de réfléchir dans un second temps au meilleur moyen de l'adapter à l'utilisateur, le « user experience design » (UX), ou encore « user-centered design » part de l'utilisateur et s'attache à concevoir les interactions les mieux à même de lui permettre d'obtenir un certain type de résultat ; l'artefact vient dans un

second temps comme le dispositif donnant lieu à ces interactions. Le design interactif ne s'identifie pas au UX, car celui-ci comprend aussi des secteurs de recherche consacrés à l'identification des utilisateurs potentiels, aux situations ou aux rôles qu'ils occupent, à leur comportement face au dispositif proposé, etc., mais il en constitue la composante centrale, et celle qui est le plus directement tributaire des sciences cognitives. En effet, le « pape », ou du moins une figure tutélaire du design interactif n'est autre que Donald Norman, ingénieur et chercheur inclassable, à la frontière de la psychologie cognitive et de la psychologie sociale, et l'un des fondateurs du Département de sciences cognitives de l'Université de Californie à San Diego (UCSD). Norman a repris au grand psychologue J. J. Gibson la notion d'« affordance » — qui désigne, *grosso modo*, les opportunités qu'offre à un animal (humain ou non) l'environnement dans lequel il est plongé, compte tenu de ses capacités et de ses besoins. Une porte se présente d'emblée (et non au terme d'un enchaînement d'étapes telles que « reconnaissance », « inférence », « plan d'action »...) comme « me permettant d'entrer », un arbre comme « me permettant d'échapper au prédateur qui me poursuit », etc. Sous le nom de *usability*, Norman étend l'affordance aux artefacts. L'« usabilité », c'est ce qui permet à l'utilisateur de faire coïncider l'affordance de l'objet avec sa fonction ; elle désigne donc la facilité d'utilisation, mais aussi la susceptibilité d'un artefact à se laisser appréhender, et à permettre un apprentissage rapide de son utilisation. Les « digital natives » qui nous lisent ne lisent jamais les instructions de la dernière console, smartphone ou TV : ils s'attendent à ce que ces appareils se présentent à eux en sorte qu'ils découvrent d'eux-mêmes, facilement, comment ils marchent. Cela dépend naturellement de leurs attentes, façonnées par des expériences antérieures et certaines connaissances — attentes, expériences et connaissances qui ne sont pas celles de leurs grands-parents : c'est pourquoi toute entreprise d'UX repose sur une connaissance de l'utilisateur : c'est ce qu'on appelle, dans le jargon du design et de l'IHM, le *facteur humain*, mais il prend, grâce à cette importation par Norman de la notion d'affordance, un sens précis et opératoire : le principe fondamental est d'aligner les affordances perceptibles sur les fonctionnalités visées, et d'éviter d'introduire des affordances trompeuses. Il s'agit à la fois d'exploiter les « réflexes » de l'utilisateur pour le conduire de manière rapide et fluide vers les interactions qu'il souhaite, et de ne pas l'induire en erreur en lui cachant des possibilités d'interaction ou en suggérant qui n'existent pas (« faux boutons »).

e. Les sciences cognitives appliquées au design et à l'architecture d'espace

Ce n'est pas seulement dans la conception d'objets et d'interfaces que les sciences cognitives viennent à l'appui du design. Un exemple caractéristique est fourni par l'étude des bases cognitives et neurales de la mémoire spatiale et de la perception multi-sensorielle, pour des applications liées à l'orientation, à l'apprentissage, à la navigation dans des environnements complexes : immeubles, gares et aéroports, centrales nucléaires, stations spatiales, environnements virtuels 3D navigables... Il s'agit, d'abord, de concevoir des cartes pour guider l'exploration dans ces espaces — ce qui est une première forme de design interactif. Mais cela va jusqu'à concevoir des espaces (ou de la signalétique), à partir d'une meilleure compréhension de la représentation cérébrale de l'espace : on part alors de l'utilisateur pour façonner un environnement spatial qui sera pour ainsi dire « à sa main ». Les termes de neuro-architecture et de neuro-design ont donc fait leur entrée récente dans le vocabulaire des sciences cognitives appliquées, tout comme la thématique du « *wayfinding* » (Hegarty 2011). Cette dernière se place, encore une fois, à la croisée entre recherche fondamentale en cognition spatiale et recherche appliquée, pouvant informer des choix pratiques.

f. Analytique visuelle

Comprendre comment les dispositifs visuels (images, diagrammes) interagissent avec les systèmes visuo-spatial et attentionnel, avec la mémoire, avec le langage, avec différentes formes de raisonnement, permet d'anticiper et de guider le design d'interfaces, par exemple comme outil éducatif ou dans le contexte de la recherche scientifique, mais aussi, de manière très générale, pour toute tâche de navigation et de recherche d'informations. Les retombées de ces recherches - alimentées par le développement des interfaces graphiques et par leur diffusion massive - investissent différents champs du quotidien, et une grande variété de champs professionnels. Un champ de recherche multidisciplinaire connexe à celui de l'IHM s'est développé dans les années 2000, sous le nom de *Visual Analytics* (VA), parfois traduit en « analytique visuelle » (Sun *et al.* 2013). Ce champ repose sur une alliance entre expertise en cognition humaine, analyse numérique automatisée des données et visualisation informatique, permettant de transformer des données en stimuli qui peuvent être traités visuellement et avec lesquels l'utilisateur peut interagir à différents niveaux. En ce qui concerne les aspects cognitifs, le VA ne se limite pas à prendre en compte les compétences perceptives visuelles humaines, mais puise dans la recherche sur la psychologie du jugement et de la décision en situation incertaine (et ne se réduit donc pas au domaine plus ancien de la visualisation de l'information). Son but est de faciliter, via la conception d'interfaces graphiques adaptées, l'utilisation et la compréhension de l'information pour guider la

prise de décision. En pratique, le champ d'application de l'analytique visuelle s'étend à toutes ces situations dans lesquelles la charge informationnelle rend impraticable une analyse rapide pour aboutir à une prise de décision.

Dans le domaine de la santé, par exemple, la possibilité de collecter et conserver de grandes quantités de données concernant chaque patient crée à la fois des opportunités et des difficultés d'exploitation de l'information disponible ; ces difficultés touchent à la fois le clinicien, l'administrateur et le chercheur. Tous s'intéressent, d'un point de vue différent, aux populations de patients. L'immense quantité de données, et la difficulté de les synthétiser, peut donc poser problème aussi bien dans la prise de décision médicale que dans la recherche.

La santé n'est pas le seul domaine dans lequel la possibilité de collecter et d'analyser une quantité de données qui dépasse les capacités humaines rend nécessaire le développement de solutions hybrides, où l'informatique et les sciences cognitives viennent en aide à la prise de décision. L'analytique visuelle s'applique aussi bien à des questions de sécurité et d'anti-terrorisme qu'à l'anticipation de catastrophes naturelles, l'analyse de trajectoires, ou encore à l'analyse des préférences des consommateurs.

2.2. Le monde augmenté

a. Réalité virtuelle, interfaces multi-modales

Entre la fin des années 1970 et les années 1990 se développe un nouveau type d'interface homme-machine, dont la caractéristique principale est de rendre possible des interactions plus multimodales, incluant la motricité. La réalité virtuelle ou augmentée se caractérise par la transmission de stimuli visuels et sonores, mais aussi tactiles, haptiques, et kinesthésiques, grâce à l'utilisation d'interfaces capable de gérer des stimuli mécaniques. Les interfaces haptiques sont aussi dotées de senseurs, qui, captant le mouvement de l'utilisateur et le traduisant en commandes pour la machine, créent une boucle entre action et perception.

Les dispositifs de réalité virtuelle varient en réalité grandement en ce qui concerne leur niveau d'immersion et d'interactivité (qui peut être assurée aussi par d'autres types de senseurs, par exemple ceux qui enregistrent les mouvements oculaires). La notion même d'immersion n'est pas univoque. Tantôt elle indique le nombre de modalités sensorielles sollicitées par le dispositif (visuelle, auditive, tactile, kinesthésique) : l'immersion est réputée proportionnelle à ce nombre ; tantôt elle renvoie à la capacité du dispositif d'isoler l'utilisateur des stimulations de provenance du monde réel.

C'est notamment à partir des années 1990 que l'on voit apparaître les premières applications de réalité virtuelle à destination du grand public, dans les domaines du patrimoine culturel, des jeux vidéo (Eichenbaum et al. 2011), des thérapies comportementales des phobies, puis de la rééducation fonctionnelle de patients nécessitant un entraînement sensori-moteur intense, et plus généralement, des simulations immersives et réalistes dans divers situations professionnelles (par exemple, des simulateurs de vol chez les élèves pilotes ou de tour de contrôle chez les aiguilleurs du ciel).

Les sciences cognitives jouent un rôle clé dans l'étude des processus perceptifs et des boucles perception-action à l'œuvre dans l'utilisation de la réalité virtuelle (Burdea & Coffet 2003) : pour réduire les effets indésirables (par exemple la sensation de nausée et de déséquilibre provoquée par certains dispositifs visuels) ; pour la compréhension de phénomènes dits de « présence », c'est-à-dire l'impression qu'un personnage est présent en chair et en os, et plus généralement que l'expérience simulée est une expérience réelle. La capacité de mesurer ces processus et boucles devient un aspect crucial de l'interaction entre développement et sciences cognitives.

Un domaine spécial de la réalité virtuelle est l'interaction *via* l'ordinateur avec des agents virtuels (avatars). Se pose alors, en plus des problèmes de crédibilité de l'expérience perceptive, un problème d'acceptabilité et de crédibilité de l'interaction sociale avec un agent artificiel, souvent anthropomorphe. Même du côté de la programmation des agents virtuels, le défi n'est pas que technologique, car il implique la simulation de comportements et modalités d'interaction humaines : regard, mouvements corporels, mais aussi perception, mémoire, capacités d'adaptation et d'apprentissage – un défi que tente de relever également l'intelligence artificielle (IA) sur laquelle nous revenons brièvement plus bas.

La réalité virtuelle évoque ainsi des scénarios de science-fiction, des expériences qui étaient jusqu'à récemment confinées au domaine du dysfonctionnement cérébral ou de la métaphysique, telles que la sensation, lors de l'immersion dans des environnements irréels, de se trouver « comme en dehors de son propre corps ». Les films mettant en scène des mondes virtuels et systèmes d'augmentation cérébrale se multiplient dans les années 1980-1990, jusqu'à aboutir à l'imaginaire de *The matrix*, où la réalité se révèle être rien d'autre qu'une simulation numérique très perfectionnée, et l'espèce humaine, une collection de cerveaux dans une cuve. Si la réalité virtuelle est encore loin de ces scénarios, elle n'en suscite pas moins de réactions de

peur : ses pratiquants seraient menacés de perdre contact avec le réel, une crainte qui s'exprime déjà envers les jeux vidéo et les communautés en ligne, assortie des risques d'addiction ou d'usage pathologique, et de passage à la violence réelle en première personne.

b. Objets, surfaces, agents intelligents, réactifs, connectés : la réalité augmentée

Si la réalité virtuelle est encore souvent associée soit à des technologies de laboratoire, soit aux jeux vidéo, les « objets intelligents » promettent d'investir la vie quotidienne, la maison, la ville. Dans le cas de la réalité augmentée, le monde réel n'est pas remplacé par l'environnement artificiel : l'utilisateur reçoit des informations des deux environnements simultanément. Les objets intelligents par contre font partie de l'environnement physique, mais ont un « cœur électronique » : ils embarquent des senseurs et des logiciels qui leur permettent d'extraire des informations de l'environnement et d'agir. Ils peuvent de plus être connectés entre eux.

Les applications de ce genre de système investissent des champs extrêmement variés : du thermostat intelligent qui contrôle la température à distance, aux puces qui surveillent le taux de molécules dans le sang, aux plaquettes placées à côté des œuvres d'art dans les musées, et qui interagissent avec des logiciels embarqués dans les téléphones portables. L'impact des sciences cognitives sur ce domaine propre à l'ingénierie est multiple : premièrement, elles permettent de concevoir des interfaces qui respectent les facultés cognitives humaines ; deuxièmement, elles identifient l'impact de ces technologies sur la cognition ; troisièmement, elles mesurent les effets cognitifs indésirables de ces technologies (par exemple, la surcharge attentionnelle ou informationnelle). Enfin, il existe potentiellement un retour de ces technologies sur la recherche en sciences cognitives – y compris la recherche fondamentale. Les interfaces qui se portent (téléphones, montres, etc) peuvent en effet permettre d'enregistrer comportements et réponses physiologiques *in situ*, en situation écologique, plutôt qu'en laboratoire, et donc de collecter des quantités importantes de données jusqu'ici indisponibles au chercheur. C'est là l'une des voies par lesquelles la recherche appliquée peut alimenter la recherche fondamentale.

3. Les agents cognitifs, la cognition externalisée

Si la section précédente était concernée par les interfaces, objets et dispositifs qui complètent la relation entre l'humain et son environnement, la présente section concerne diverses tentatives pour remplacer l'agent ou pour externaliser la cognition humaine, en partie ou dans son intégralité.

Dans ce domaine, la cognition humaine (ou animale) constitue le modèle dont ces dispositifs s'inspirent de plus ou moins près. Là encore, les efforts pour créer ces systèmes influencent à leur tour la recherche fondamentale en poussant nos conceptions de la cognition humaine dans de nouvelles directions. Ces efforts sont menés sous les intitulés généraux de l'intelligence artificielle (IA — en anglais, AI pour *artificial intelligence*) et de programmes qui en dérivent mais ont acquis une autonomie plus ou moins grande : l'apprentissage automatique (plus connu sous son intitulé anglais de *machine learning* ou ML), la robotique cognitive¹, le traitement du langage naturel², la vision artificielle³.

3.1 L'intelligence artificielle

a. Quel est le rapport entre IA et sciences cognitives ?

L'IA et les sciences cognitives ont émergé au milieu des années 1950, simultanément et dans un même mouvement d'idées⁴ : le projet de l'IA à ses débuts était de produire un ordinateur intelligent, c'est-à-dire capable d'accomplir toute tâche exigeant de l'homme le recours à son « intelligence », par quoi les pionniers de l'IA entendaient à peu près ce qu'on désignerait aujourd'hui comme ses capacités cognitives.

La relation entre le projet de l'IA et celui des sciences cognitives est un point nodal sur lequel les opinions ont varié et continuent de le faire. On peut distinguer quatre positions :

1. L'identité : les deux entreprises sont en fait une seule et même chose, vue sous des angles

¹ Voir, dans le chapitre « Action », l'encadré sur la robotique.

² Encadré dans le chapitre « Langage » ?

³ Encadré dans le chapitre « Vision » ?

⁴ Voir article Andler. Les sciences cognitives n'ont reçu leur nom actuel que bien plus tard, mais le projet et les ressources pour le mettre en œuvre ont pris forme dès ce moment.

différents ; c'était l'avis de pionniers de l'IA tels que Herbert Simon ou encore Roger Schank, dont l'adage le dit bien : « L'étude de l'intelligence artificielle n'est rien d'autre que l'étude de l'intelligence naturelle. »

2. L'inclusion (stricte) de l'IA dans les sciences cognitives, pour l'une ou l'autre des deux raisons suivantes, ou les deux ensemble :
 - a. l'IA ne concerne qu'une partie des sciences cognitives, celle qui porte directement sur l'intelligence au sens courant de la capacité de résoudre des problèmes, plus ou moins difficiles, dans une variété illimitée de contextes (depuis la rédaction d'un chèque à partir du montant en chiffres jusqu'au jeu de go, en passant par les puzzles logiques, la planification d'une action complexe, etc.) ;
 - b. l'IA constitue une méthodologie parmi d'autres au sein des sciences cognitives : elle recherche des théories descriptives et explicatives de la cognition sous la forme de programmes informatiques, qu'elle teste en comparant leurs performances à celles des humains.
3. Le « cousinage »: l'IA et les sciences cognitives étudient chacune deux espèces d'un même genre, la cognition (ou l'intelligence) en tant que fonction, l'IA se consacrant à sa réalisation « sur silicium » (informatique), les sciences cognitives à sa réalisation « sur support humide », c'est-à-dire neurale. Elles ont pour intersection l'étude de la cognition comme fonction, et comportent chacune une partie propre consacrée à la réalisation, sur chacun des supports, de cette fonction.
4. L'analogie : l'IA cherchant à simuler les fonctions cognitives humaines s'inspire à l'occasion des sciences cognitives, qui lui fournit des idées, des pistes ; inversement, lorsque l'IA découvre une manière de réaliser une certaine fonction, les sciences cognitives peuvent s'en servir comme heuristique pour comprendre comment cette fonction est réalisée par le cerveau. Mais rien n'oblige à supposer que le cerveau et l'ordinateur doivent s'y prendre de la même façon pour réaliser une fonction donnée — qu'il s'agisse de gagner au jeu de go ou de reconnaître un visage. De plus, il y a différents sens en lesquels on peut dire que deux réalisations d'une fonction sont « les mêmes » : au niveau de détail le plus fin, il y a nécessairement une différence, car le cerveau est organisé, en tant que système physique, très différemment d'un ordinateur, voire d'un réseau de neurones artificiels ; à l'autre extrême la fonction est par définition la même. La question est donc de savoir s'il existe un niveau intermédiaire significatif tel qu'on puisse considérer qu'à ce niveau, les deux systèmes procèdent de la même manière, ou plutôt, puisque le système artificiel reste à concevoir, si l'on peut ou si l'on doit le concevoir en sorte de réaliser cette condition. Différentes écoles au sein de l'IA et au cours de son histoire ont proposé des réponses très différentes.

La principale conclusion à tirer, pour ce qui nous concerne, est que l'IA ne peut être considérée comme une simple aire d'application des sciences cognitives. Elle ne l'est qu'à condition de choisir la quatrième orientation, qui établit un rapport d'analogie entre IA et sciences cognitives, et de donner à la toute dernière question la réponse suivante : c'est au niveau informationnel que l'IA ainsi conçue cherche à construire des programmes qui sont isomorphes aux suites d'opérations réalisées par le cerveau pour réaliser les fonctions visées. Un exemple d'école serait le suivant : si l'on a réussi à déterminer la suite d'opérations par laquelle le système auditif humain repère une source sonore, on cherchera à construire un dispositif computationnel

b. L'IA de l'autonomie à la complémentarité

Entre le milieu des années 1950 et la fin des années 1970, l'IA a connu des débuts prometteurs et un déclin rapide ; ce déclin doit-il être compris comme un échec, ou comme un simple « trou d'air » comme en connaissent régulièrement les programmes de recherche ? Ce n'est pas le lieu d'en discuter, mais nous pouvons retenir deux choses :

- (i) En tant que programme de sciences cognitives, la méthodologie de cette première phase de l'IA a échoué : ce n'est pas en essayant de simuler des fonctions cognitives à partir des intuitions que l'on peut avoir comme informaticien ou logicien qu'on peut espérer comprendre les mécanismes cognitifs ; ce qui manque, c'est une approche *scientifique*, au sens des sciences empiriques, des processus mentaux et cérébraux, processus qui ne sont que très partiellement accessibles à l'intuition ou au bon sens.
- (ii) L'IA a tiré les conséquences de cet échec et renoncé à son ambition de créer une intelligence *générale*, révisant ses ambitions à la baisse et visant désormais à mettre au point des programmes « intelligents » au sens qui nous est désormais familier, celui en lequel une « app »,

un système de traitement de texte, une aide en ligne, un « assistant » intégré au système d'exploitation d'un smartphone ou d'un ordinateur, un système de navigation routière, un téléviseur, un ascenseur, un carburateur, un système de surveillance, un système de recommandation d'achats, une pompe à insuline, etc. etc., sont « intelligents » (les guillemets sont souvent incluses dans la désignation). L'intelligence en question est ce en vertu de quoi le logiciel est capable de tenir compte de certains paramètres contextuels, liés à l'utilisateur ou à ses activités récentes notamment. Pour le dire en un mot, l'IA s'est donné pour objectif de fabriquer de l'intelligence *ciblée*, destinée à soulager l'agent humain en effectuant à sa place un certain nombre de réglages, ou en automatisant une série d'opérations fréquemment utiles dans certains circonstances.

De manière générale, l'IA, du moins jusqu'à tout récemment, a renoncé à créer des dispositifs jouissant d'une intelligence *autonome* : elle a visé au contraire à établir la meilleure *complémentarité* entre l'agent humain et le système informatique. Elle se fonde, pour y parvenir, sur deux types d'approche :

- l'approche analytique, qui consiste à décomposer une fonction, une situation, un plan d'action en segments plus simples, et d'automatiser ceux qui s'y prêtent ;
- l'importation directe de compétences humaines, consistant à obtenir d'individus humains des réponses à une série de questions relatives à un domaine de compétence donné, sans chercher à savoir comment ces réponses sont obtenues par les experts interrogés.

Les deux approches sont complémentaires. Si l'approche analytique avait suffi, l'IA des commencements aurait été sur la bonne voie. C'est quand elle s'est aperçue des impasses auxquelles elle était conduite qu'elle a changé de stratégie. Ce fut l'époque des *systèmes experts*. Puisqu'on ne pouvait décomposer analytiquement les compétences nécessaires à l'action intelligente dans un domaine donné, on irait accoucher des experts humains de leur savoir-faire, et on consignerait leurs maximes dans une banque de données, sous la forme de « règles de production » de la forme : SI condition1, ALORS action1, où action1 peut consister soit à effectuer ou programmer un comportement donné (un geste, une inscription...), soit à déclencher une seconde règle de la forme SI condition2, ALORS action2, et ceci jusqu'à ce que le problème soit résolu : identification du cas examiné, détermination de la mesure à prendre, etc.

Les systèmes experts, après avoir donné quelques bons résultats, ont vu leur domaine d'application fortement restreint ; on les trouve aujourd'hui intégrés à certains systèmes d'aide à la décision (*decision support systems* ou DSS en anglais), notamment dans le domaine médical. Mais l'idée de recourir aux capacités humaines pour fournir à l'ordinateur les informations complexes qu'il ne peut acquérir de lui-même, du moins en un temps raisonnable, est largement utilisée aujourd'hui. À titre d'exemple, la reconnaissance visuelle d'images a fait ces dernières années de remarquables progrès grâce à l'accroissement de la mémoire et de la vitesse de traitement, mais aussi de l'existence d'immenses bases d'images ; encore faut-il que ces images soient préalablement catégorisées par des humains. C'est ainsi que le récent projet ImageNet a emmagasiné un milliard d'images, qui ont été étiquetées et ramenées à 22000 catégories grâce au travail de près de 50000 personnes mobilisées par le système Mechanical Turk d'Amazon.

c. En pratique, comment l'IA se situe-t-elle aujourd'hui ?

L'IA, après une traversée du désert, est aujourd'hui un domaine en pleine expansion qui relève essentiellement de la technologie et de l'ingénierie, même s'il inclut une composante théorique. Il est hors de question de présenter ici les nombreux sous-domaines qui la composent, et qui entretiennent avec les sciences cognitives des rapports très différents et en rapide évolution (Russell & Norvig 2009). Nous nous bornerons à évoquer cinq orientations générales au sein de l'IA d'aujourd'hui.

- (i) Une bonne partie des réalisations de l'IA consiste en systèmes hybrides destinés à seconder l'agent humain dans des tâches particulières, en s'appuyant généralement sur la « mise en conserve » de compétences humaines récoltées auprès d'experts ou à partir de situations réelles dont l'issue est connue ; ce secteur de l'IA est presque entièrement immergé dans l'informatique — de fait, ce genre d'AI peut être assimilé à une forme d'informatique avancée ou spéciale.
- (ii) Certaines branches de l'IA ont acquis une relative autonomie, correspondant à un stade plus avancé de théorisation, rendu possible par une compréhension plus complète du phénomène naturel correspondant. Il s'agit de la vision artificielle (*machine vision* en anglais), du traitement automatique du langage naturel (TAL ou TALN, en anglais *natural language processing* ou NLP),

et de l'apprentissage automatique, ou artificiel (AA, en anglais *machine learning* ou ML). Nous l'avons déjà signalé, et y revenons ci-dessous, en notant ici que l'apprentissage automatique joue au sein de l'IA un rôle très spécial.

- (iii) Les progrès récents dans ces trois derniers domaines suscitent chez certains l'espoir et/ou la crainte que l'IA sera bientôt en mesure de créer des dispositifs totalement autonomes, depuis les véhicules autopilotés jusqu'aux guerriers robots, voire à des agents artificiels super-intelligents prenant les décisions à la place des individus et à la limite le contrôle des affaires humaines. Ces perspectives sont néanmoins éloignées et peut-être infiniment éloignées.
- (iv) Dans certains domaines, principalement mais pas seulement celui de la vision artificielle, les chercheurs commencent à puiser une inspiration dans les architectures corticales, dans la manière dont le cerveau « calcule » pour réaliser certaines fonctions. Ici le rapport aux sciences cognitives est sans conteste celui de l'application : ce que les neuroscientifiques et les mathématiciens, très actifs dans le domaine, nous apprennent sur le cortex visuel sert d'heuristique pour la vision artificielle.
- (v) Une branche récente et encore marginale veut renouer avec l'ambition initiale de l'IA : construire non des logiciels spécialisés, dont l'utilité est strictement limitée à une tâche donnée, mais une intelligence générale, au sens où l'est, du moins en apparence ou à un certain niveau de description, l'intelligence humaine, capable de passer de manière fluide d'un mode à un autre, d'une situation à une autre, sans faire rentrer préalablement chaque contexte, chaque question, dans une case prédéfinie pour lui appliquer un algorithme spécialisé. Le label que s'est donné ce mouvement est *intelligence générale artificielle* (en anglais *artificial general intelligence (AGI)*). Les fondements et l'avenir de cette tentative (qui rejoint et amplifie le mouvement d'opinion indiqué en (iii)) sont des questions controversées, mais il est notable que parmi les atouts que les tenants de l'AGI estiment détenir figure précisément le fait de pouvoir s'appuyer sur des sciences cognitives développées, alors qu'elles étaient balbutiantes à la naissance de l'IA.

3.2 L'apprentissage automatique

a. L'apprentissage en général et la place de l'apprentissage automatique

Apprendre, dans le langage courant, s'entend de plusieurs façons. L'apprentissage automatique (AA) s'intéresse notamment à deux formes simples : apprendre des faits particuliers, et acquérir la capacité de rattacher des objets d'un certain ensemble à une catégorie au sein de cet ensemble (par exemple, les automobiles circulant sur les routes à leur marque). L'apprentissage d'une langue, d'une capacité motrice, d'un savoir-faire complexe sont des formes plus évoluées que l'AA n'a pas encore abordées.

Apprendre des faits, les grandes dates de l'histoire de France, le nom des capitales africaines ou le prénom de ses étudiants, et de manière générale acquérir des connaissances factuelles particulières, est un premier type d'apprentissage, courant mais d'intérêt théorique limité : ce qu'on apprend ne s'étend pas au-delà de la base d'apprentissage, on est seulement devenu capable de reconnaître, se remémorer, répéter les faits appris. Cependant, apprendre beaucoup de faits peut permettre d'acquérir une méthode et d'améliorer une capacité plus large, celle d'acquérir des connaissances factuelles quelles qu'elles soient, ou d'améliorer cette capacité : apprendre plus vite, plus sûrement, plus facilement... On touche ainsi à un sens plus intéressant d'apprendre : acquérir de manière durable une nouvelle capacité, ou améliorer une capacité déjà présente — apprendre à faire quelque chose, ou à le faire mieux, par opposition à apprendre *que* quelque chose porte tel nom, ou s'est produit à telle date, etc. Le cas particulier sur lequel porte, pour la plus grande part, l'apprentissage automatique, est celui dans lequel la capacité acquise est de ranger des items dans des catégories prédéterminées : tel tableau est un Picasso, tel autre un Braque ; celui-ci est un faux Picasso, celui-là un Picasso authentique ; ceci est un signal sonar de mine (DANGER), cela un signal de rocher (SANS DANGER) ; ceci est un signal « stop » (S'ARRÊTER), cela un signal « voie prioritaire » (NE PAS S'ARRÊTER) ; ceci est un chien agressif (SE MÉFIER), cela est un chien pacifique (PASSER TRANQUILLEMENT), etc.

Cependant, il existe entre les deux types d'apprentissage dont nous partons, et entre ces types et d'autres apparemment plus intéressants, des rapports étroits. Dans l'ordre inverse, remarquons d'abord que de nombreuses capacités peuvent se ramener, moyennant certaines interfaces, à une aptitude à catégoriser : savoir conduire peut peut-être se décomposer en la possession de certaines routines, acquises une fois pour toutes, et la capacité de catégoriser les situations rencontrées sur la route sur la base d'indices sensoriels ; savoir survivre dans un quartier dangereux repose au moins en partie sur la capacité d'identifier correctement les situations, rues, bars, individus... comme étant potentiellement dangereux ou probablement sûrs, et ainsi de suite.

Pour ce qui est des rapports entre l'acquisition de faits singuliers et l'acquisition d'une capacité à catégoriser, ils sont depuis longtemps théorisés sous le nom d'apprentissage inductif. Si l'on me présente 100 prénoms masculins et 100 prénoms féminins courants en France entre 1900 et 1914, en me précisant à chaque fois le genre du prénom présenté, j'acquerrai probablement la capacité de reconnaître le genre d'un nouveau prénom de la même période. De même si l'on me présente 100 relevés sonars de mine et 100 relevés sonars de rocher, je pourrai, avec une certaine chance de réussite, identifier un nouveau relevé. Il se peut aussi que 200 relevés d'entraînement ne suffisent pas, qu'il en faille 1000 ou 5000... ; il se peut même qu'aucun échantillon, si grand soit-il, ne permette pas d'acquérir la moindre compétence en la matière : le problème inductif est alors sans solution, mais beaucoup de problèmes inductifs qu'on rencontre dans la pratique sont solubles.

L'apprentissage automatique désigne, de manière très générale, l'ensemble des processus par lesquels un ordinateur, ou plus largement un système numérique de traitement de l'information, améliore ses performances. Un cas particulier, en pratique le cas de loin le plus important numériquement et sur le plan des applications, est celui de l'apprentissage inductif ; de fait, dans beaucoup de présentations, l'apprentissage automatique est *par définition* l'ensemble des techniques permettant à un système informatique d'acquérir la capacité de catégoriser un certain type d'objet, sur la base d'un répertoire d'exemples. Nous nous concentrerons donc sur ce cas central.

b. Types et systèmes d'apprentissage automatique

Ce que nous avons présenté à l'instant est le cas le plus simple, celui de l'apprentissage dit *supervisé*, dans lequel l'apprenant est informé de la catégorie correcte de l'item présenté. Il existe deux autres cas, celui de l'apprentissage non supervisé, et celui de l'apprentissage par renforcement. Dans le premier, l'apprenant constitue lui-même les catégories pertinentes, sur la base d'hypothèses très générales ; il y parvient en repérant certaines régularités statistiques dans les données présentées. Dans le second, l'apprenant ne reçoit que des informations partielles sur la qualité de sa catégorisation : il peut apprendre qu'elle est incorrecte, sans plus ; il peut aussi apprendre que la suite de ses conjectures est plus ou moins bonne, sans savoir quels éléments sont corrects et lesquels sont faux. Aujourd'hui, les systèmes d'AA sont très souvent hybrides, combinant deux des approches, voire les trois (Flach, 2012).

Par ailleurs, quel que soit le cadre général, un certain nombre de critères s'appliquent. Ils concernent notamment le nombre d'exemples requis, la structure statistique de l'ensemble d'exemples, la vitesse de convergence vers un système de catégorisation acceptable, en fonction de la probabilité et du genre d'erreurs tolérés, la complexité algorithmique de l'apprentissage...

Sur le plan opératoire, l'AA a connu ces dernières années une accélération considérable, grâce au succès d'une approche, appelée *deep learning* (on dit aussi en français, moins couramment, *apprentissage profond*), associant une certaine architecture (des réseaux multicouches de neurones formels) et un algorithme d'apprentissage inspiré de la physique statistique (Goodfellow *et al.* 2016). Ces éléments étaient en place dès le milieu des années 1980, mais il a fallu, pour qu'ils conduisent aux succès actuels, la conjonction des énormes capacités computationnelles d'aujourd'hui, notamment par le biais de la mise en réseau de très nombreuses machines, la constitution d'immenses bases d'exemple (en particulier d'images), des perfectionnements mathématiques et algorithmiques, et la possibilité d'essayer en parallèle de nombreuses variantes et de voir lesquelles fournissent les meilleurs résultats. Le deep learning s'applique aujourd'hui à une grande variété de situations, allant des véhicules autopilotés à la gestion de la consommation électrique de méga-entreprises ou au diagnostic médical.

c. Apprentissage automatique, intelligence artificielle et sciences cognitives

L'AA est constituée pour une bonne part de techniques formelles avancées : au niveau théorique il s'agit d'un domaine de science appliquée hautement spécialisé, à la frontière de plusieurs spécialités relevant des mathématiques appliquées et de l'informatique théorique. Mais, comme on l'a indiqué, l'AA comprend aussi un questionnement plus large, concernant la nature et les mécanismes de l'apprentissage en général, non limité à la catégorisation inductive à partir d'exemples, étiquetés (apprentissage supervisé) ou non. Dès lors, les sciences cognitives rentrent dans le tableau, à la fois comme source et comme cible d'idées.

Les sciences cognitives peuvent éclairer des formes d'apprentissage propres aux humains ou à d'autres espèces animales, dont l'AA ne peut aujourd'hui imaginer d'équivalents : les humains apprennent constamment, et pas seulement à catégoriser des entités d'un certain type ; ils acquièrent des savoir-faire, des connaissances, dont certains sont spécifiques à un domaine (servir au tennis), mais d'autres semblent pertinentes dans de nombreux domaines (raisonner), et se combinent à d'autres apprentissages (apprendre à compter, apprendre les relations sociales dans une école, apprendre ce qu'est une discipline...). Le cerveau

apprend également, à son niveau, selon des procédures qui pourraient être reproduites dans des systèmes artificiels. Inversement, les architectures les plus efficaces découvertes en AA peuvent inspirer les sciences cognitives, ainsi du reste que certains concepts et résultats généraux des théories computationnelles de l'apprentissage développées dans le cadre de l'AA.

Quant au rapport entre AA et IA, il pose deux questions, l'une facile, l'autre beaucoup moins. La question facile est de savoir si l'AA fait partie de l'IA ; la question difficile de savoir si elle en est une partie essentielle. Pour la première, dès lors que l'IA contemporaine est revenue sur les hypothèses princeps de ses débuts, celles du cognitivisme symbolique (v. le chapitre « Les sciences cognitives : d'où viennent-elles... »), et qu'elle incorpore toute technologie susceptible de se substituer à l'agent humain dans l'exécution d'une fonction cognitive, la réponse est clairement positive. Pour la seconde, la difficulté est la suivante. Nous avons vu, dans un passage qui a pu paraître un peu abstrait (§3.1.a), qu'il existe différentes conceptions de ce qui constitue un modèle d'intelligence artificielle, selon le « grain » ou l'échelle à laquelle le modèle ressemble à l'intelligence humaine, ou à une de ses composantes. L'IA des débuts exigeait un grain assez fin : un modèle d'IA de résolution du problème des tours de Hanoï, par exemple, devait effectuer une série d'étapes clairement identifiables comme des moments d'un raisonnement humain possible. Un système computationnel qui obtient les résultats voulus sans qu'on puisse lui attribuer une série d'états et de transformations internes correspondant à quelque chose de reconnaissable dans la cognition humaine ne satisfait pas ce réquisit. On peut cependant le considérer comme une contribution à l'IA, à titre d'étape préparatoire : après tout, les états et transformations internes du cerveau sont loin de pouvoir être mises, au stade présent des connaissances, en correspondance avec des entités relevant de notre flux de pensée. Mais, pour qui maintient l'exigence d'une correspondance intelligible entre états mentaux conscients et trajectoire du dispositif artificiel, celui-ci reste en quelque sorte en marge de l'IA. Or c'est précisément le cas des modèles du *deep learning*, paradigme de l'AA. Dans une telle perspective, donc, l'AA reste à la périphérie de l'IA. À l'autre extrême se situent deux positions, qui peuvent ou non être combinées. L'une est que, pour l'IA, seul compte le résultat — si un tomahawk peut être utilisé comme système cognitif, sans qu'on aie aucune idée de la manière dont il produit ce « comportement » intelligent, alors le tomahawk est au cœur de l'IA, et peut même complètement l'absorber s'il parvient à une forme complète d'intelligence. Selon la seconde position, nous nous trompons en prenant pour niveau fondamental les étapes de la pensée consciente en mode de résolution de problème : il est possible — certains le pensent — que notre intelligence soit en réalité basée sur nos capacités d'apprentissage inductif, et si tel était le cas, l'AA fournirait bel et bien, au sens littéral, un modèle de la cognition humaine, et à ce titre se confondrait avec l'IA.

3.3 La robotique cognitive

Contrairement aux domaines dont il vient d'être question, la robotique cognitive n'est pas un champ disciplinaire clairement délimité. Il désigne plutôt une orientation au sein de la robotique, caractérisée par :

- (1) la nouveauté : elle se conçoit comme une avant-garde de la robotique classique ;
- (2) le projet de conférer aux robots des aptitudes cognitives autonomes, leur permettant en particulier, face à des situations non prévues, de comprendre ce qui se passe, de raisonner, de prendre des décisions, de former des plans d'action et de les mener à bien (on parle parfois à ce propos de *robotique intelligente*) ;
- (3) le choix méthodologique de recourir le plus largement possible aux sciences cognitives, et d'y contribuer en retour ;
- (4) et dans cette dernière perspective, plus spécifiquement, de donner un contenu scientifique et technologique précis à l'idée de la cognition incarnée (*embodied* en anglais) et impliquée (*embedded*), ou encore « ancrée » (*grounded*), selon l'appellation la plus récente, qui combine les deux aspects, et selon laquelle la cognition n'est pas la propriété d'un organe (le système nerveux) ou d'un système de traitement de l'information (un ordinateur, un réseau de neurones formels, etc.) isolés et reliés à l'environnement par des capteurs et des effecteurs cognitivement inertes, mais la propriété d'un organisme, ou d'un système, possédant un corps et impliqué dans des interactions signifiantes avec l'environnement ; en particulier, selon cette conception, la perception et l'action ne sont pas deux phases ou fonctions séparées, elles sont au contraire intriquées et ne laissent donc pas entre elles une place réservée à ce qui serait la cognition proprement dite.

De manière générale, on peut distinguer une robotique cognitive modérée ou pragmatique, et une robotique cognitive ambitieuse ou visionnaire. La première cherche à conférer à ses robots des capacités cognitives plus évoluées, par tous les moyens possibles, qu'ils proviennent ou non des sciences cognitives, des

neurosciences voire de la biologie évolutive et de la vie artificielle (VA). La seconde a beaucoup en commun avec la première, mais vise à constituer un cadre nouveau, subsumant à la fois la robotique et les sciences cognitives, dans lequel développer une théorie complète de l'humain abstrait, « corps » et « esprit » compris, théorie qui s'incarne dans des robots qui sont des « personnes » ou des « agents » complets, munis de tout le répertoire requis, comme elle s'incarne dans les humains. C'est ainsi que se développe une « robotique développementale » qui s'intéresse par exemple au « bébé robot », débutant dans l'« existence » avec un bagage moteur très sommaire, à l'image du nouveau-né humain, et apprenant de lui-même à actionner ses membres de manière à agir et se déplacer selon ses désirs ou ses besoins. Dans le même esprit, on s'intéresse à ce qui peut tenir lieu, chez un robot, d'autonomie, de motivation, d'émotions, de conscience, avec un triple objectif : concevoir des robots capables de se substituer aux humains dans un répertoire de tâches et de situations aussi large que possible, ce qui implique qu'ils puissent à la fois comprendre et éprouver ces états ; développer une déontologie adaptée à ces robots (déterminer s'ils sont des droits et des devoirs, et si nous en avons à leur égard) ; jeter une lumière nouvelle sur ces phénomènes chez l'homme, sachant qu'ils sont loin d'être encore bien compris.

3.4. Le traitement automatique du langage naturel

Avant même que l'IA ne prenne forme et adopte le nom qu'elle porte aujourd'hui a été lancé, principalement aux Etats-Unis, le projet de traduction automatique (*machine translation* en anglais) — la motivation étant de pouvoir assimiler rapidement l'information scientifique contenue dans les revues publiant en russe, afin notamment de rattraper un retard, réel ou supposé, en matière technologique. Le projet, après quelques succès initiaux, s'est rapidement heurté à des obstacles insurmontables, cet échec étant l'un des facteurs qui a conduit les chercheurs à formuler le projet plus global d'intelligence artificielle, dans le cadre duquel la question de la traduction automatique s'est trouvée plongée dans une problématique plus vaste, celle du traitement [automatique] du langage naturel (TAL ou TALN, *natural language processing* ou NLP en anglais).

Ce domaine est à la fois vaste et d'une importance énorme. Il est vaste parce que le langage est étroitement lié à la cognition humaine, tout particulièrement dans le monde d'aujourd'hui, en sorte qu'une très forte proportion de nos états mentaux, de nos décisions, de nos actions, font intervenir le langage, comme input et/ou comme output. Il est important parce que la seule manière naturelle de communiquer et d'acquérir rapidement des connaissances relativement complexes nécessite de partager une même langue, sous forme orale ou signée (au sens du langage sourd) et sous forme écrite. Or cette situation de partage n'est réalisée qu'entre individus humains parlant la même langue. Elle ne l'est pas dans deux familles de cas. La première est celle dans laquelle les individus concernés parlent des langues différentes : un traducteur est nécessaire, et s'il n'y en a pas de disponible, on est ramené au second genre de cas, dans lequel l'une des parties en présence n'est pas un humain mais un système artificiel de traitement de l'information, ce qui à l'ère du numérique se produit constamment. Sans capacité de traiter le langage des humains, la machine est lourdement handicapée et son utilité fortement réduite : elle doit pouvoir comprendre ce que dit ou écrit un humain, et elle doit pouvoir produire des messages qu'à son tour il puisse comprendre.

Mais que signifie « comprendre » ? Cette question relève de la philosophie du langage et de l'esprit, et elle n'a pas à ce jour reçu de réponse qui fasse l'unanimité. Mais cela n'empêche nullement le TAL de progresser, en trouvant des façons pertinentes d'opérationnaliser la notion de compréhension, selon la tâche à accomplir.

Que sont donc, de manière plus précise, ces tâches ? Nous ne pouvons faire mieux que d'en fournir un échantillon.

On doit d'abord distinguer le traitement du langage écrit, du langage oral, et le passage de l'un à l'autre. Chacune de ces modalités pose des problèmes spécifiques (pour le langage oral, par exemple, l'analyse du flux de parole et la compréhension d'un texte généralement incomplet, syntaxiquement incorrect, comportant des répétitions... ; et la production de paroles intelligibles, dans une tonalité acceptable).

En deuxième lieu figurent des problèmes qui sont indépendants de la modalité. Une première série concerne la compréhension. Un exemple simple en est l'identification de la langue particulière employée. Nettement plus difficile, l'analyse syntaxique et sémantique de la phrase, ce qui suppose en particulier la capacité de lever les ambiguïtés, de résoudre les anaphores et de manière générale de compléter les informations codées dans le texte à l'aide du contexte, de connaissances générales sur le monde et de connaissances particulières sur le domaine dont relève le texte. On peut aussi vouloir identifier des thèmes, des faits, des entités mentionnées dans le texte, et les relations qui existent entre elles ; on peut chercher à vérifier la cohérence sémantique, narrative, pragmatique du texte (par exemple pour repérer des messages codés), s'intéresser à la nature du texte (une circulaire administrative, un appel à l'aide, une invitation, un poème, un

ordre de bataille...), à sa teneur générale, à son actualité. On touche là à des problèmes qu'on ne sait résoudre que très partiellement, toujours dans des contextes particuliers.

Une deuxième série de questions concerne la production de texte, écrit ou oral, à partir de l'équivalent, pour un système artificiel, d'une intention communicative : informer l'interlocuteur d'un état de fait, répondre à sa question, l'interroger... Les réalisations en la matière sont plus limitées encore que pour la compréhension, même s'il est relativement facile de donner l'illusion d'une production riche de discours, comme le faisait le célèbre logiciel « psychanalyste » ELIZA créé au MIT par Joseph Weizenbaum dans les années 1960, et comme le font aujourd'hui (de manière moins spectaculaire mais plus honnête : personne ne s'y laisse prendre, malgré la jolie voix !) les services d'aide à la clientèle.

Toutes les fonctions dont il vient d'être question interviennent dans d'innombrables applications — indexation, résumé, compilation, extraction d'informations de textes courants, de textes littéraires ou politiques, de documents d'archive, consultation de bases de données, repérage de situations exigeant une intervention ou une surveillance, dialogue avec le consommateur, la personne en danger, le patient, la personne en situation de handicap, le technicien, le policier, le militaire, le pompier ou le médecin intervenant sur place, etc., sans oublier la traduction.

Le TAL se situe donc clairement à l'intersection de plusieurs domaines, qui sont eux-mêmes liés entre eux : les sciences cognitives, l'IA, l'informatique et, bien entendu, la linguistique (voir l'encadré dans le chapitre « Langage »). Ce qu'on a dit de la notion, encore mal élucidée, de « compréhension », s'applique à l'entreprise dans son ensemble : ses objectifs ne semblent pouvoir être atteints que moyennant la résolution de problèmes fondamentaux dans les sciences cognitives. Entre TAL et sciences cognitives, la proximité est très grande. Elle est aussi très grande avec l'IA : si l'on fait abstraction de la perception et de la motricité (une simplification que certains considèrent cependant comme gravement trompeuse, mais nous ne pouvons nous y attarder ici), pour ne garder que ses dimensions conceptuelles et inférentielles, l'IA et le TAL semblent complètement solidaires : tout progrès dans l'un nécessite et/ou entraîne un progrès dans l'autre. On pourrait donc penser que les progrès en TAL sont strictement tributaires des sciences cognitives et plus particulièrement de la linguistique, et strictement corrélés avec les progrès en IA. Or, les développements récents ne confirment pas la première hypothèse, et jettent une lumière inattendue sur la seconde.

En effet, le TAL est entré dans l'ère des statistiques, ou plutôt, rentré, car les premières tentatives, en matière de traduction automatique notamment, recouraient déjà à l'outil statistique : les co-occurrences étaient la clé pour « comprendre » un texte et en tirer l'information qu'on cherche. Ces premières tentatives avaient tôt montré leurs limites, et l'approche par les principes, fortement influencée par la linguistique théorique, s'était, raisonnablement, imposée. L'irruption du *deep learning* a profondément changé les choses : les logiciels classiques sont surclassés par la force brute de ces systèmes d'apprentissage automatique. Mais c'est précisément ce qui se passe en IA de manière générale. Le *deep learning* fait-il progresser l'IA, le TAL, la théorie computationnelle des jeux d'échecs et de go, ou bien ne fait-il que nous éviter la peine de le faire ? Il y a là une vraie question, qui sera peut-être résolue dans les prochaines années.

3.5 La vision artificielle

Le domaine de la vision artificielle est à certains égards assez semblable à celui du traitement automatique du langage. Le point de départ était concret et relativement modeste, du moins en apparence : il s'agissait de la « reconnaissance des formes » (*pattern recognition*), à des fins utilitaires : fabrication en série, détection de défauts, guidage de bras robotiques, analyse d'images (médicales, aériennes, etc.). Le problème s'est révélé plus ardu qu'on ne le pensait, et connecté à des questions plus vastes d'intelligence artificielle. La question de l'apprentissage est rapidement devenue centrale, et réciproquement la reconnaissance de formes est rapidement devenue un terrain d'expérience pour l'apprentissage automatique, et une occasion pour l'approche connexionniste (par réseaux de neurones formels) de prendre, au cours des années 1980, un avantage décisif sur les méthodes symboliques de l'IA classique. L'aboutissement de ce processus aujourd'hui rappelle également ce qui se passe en TAL : les réseaux du *deep learning*, passant en quelque sorte à saute-mouton par dessus les méthodes raffinées mises au point dans la période précédente, obtiennent des résultats excellents. Nous verrons cependant dans un court moment que ces modèles ne sont pas sans défaut.

Du côté des sciences cognitives, le rapport est également très étroit : la vision a donné lieu à une conceptualisation (due à David Marr) particulièrement précise et rigoureuse de ce qu'est une théorie computationnelle d'un phénomène cognitif, et les progrès des sciences de la vision, incluant les neurosciences, la psychologie et la psychophysique et la modélisation, n'ont pas d'équivalent dans d'autres domaines, notamment grâce à l'apport de théories mathématiques puissantes. Au total, comme dans le TAL, mais de manière plus développée, les idées circulent entre sciences cognitives, IA et vision artificielle, les modèles de *deep learning* constituant en quelque sorte un raccourci en matière de reconnaissance d'objets dans les bases

d'images bidimensionnelles, et, de plus en plus, dans d'autres situations telle que les trajets des véhicules auto-pilotés.

Mais le domaine ne se limite pas à cela. On parle en anglais de *machine vision*, ainsi que de *computer vision* — les deux termes ne sont pas clairement différenciés, le premier tendant néanmoins à désigner un sous-domaine du second centré sur les applications industrielles, le second s'intéressant aussi à des questions plus théoriques et à une série d'applications plus ambitieuses, telles que la reconnaissance de scène, la reconstruction d'un objet en trois dimensions, le design d'objets et d'espaces, la navigation... Dans sa définition la plus large, la vision artificielle veut conférer aux ordinateurs et aux robots la capacité de voir — mais que signifie voir ? Comme dans le cas du langage et de la notion de compréhension, la question est fondamentale et reste ouverte. Mais ici aussi on peut opérationnaliser la notion de différentes manières. Reconnaître un objet dans une photo, identifier un visage, repérer une tumeur sur un scan ou une mycose sur un plant de tomates constitue un premier niveau. Analyser une scène complexe est plus compliqué : il faut pouvoir la segmenter, en repérant les bords, les discontinuités et les continuités, vraies et apparentes, conjecturer les formes occluses, apprécier les distances relatives — les difficultés sont considérables et sont loin d'être entièrement surmontées, mais ici encore, beaucoup dépend de la manière dont on entend opérationnaliser la notion d'analyse : qu'attend-on du système ? De repérer une différence entre des clichés pris à deux dates différentes, par exemple la construction d'une base militaire, ou d'une route en Amazonie, d'une zone récemment défrichée, d'un changement de récolte. Ou bien, plus difficile, de reconnaître une situation : un mouvement de troupes, la propagation d'un incendie, la menace d'une collision... Ces tâches sont à la limite des capacités actuelles : certaines sont accomplies avec succès, d'autres non. Mais il s'agit toujours de reconnaître un objet, un état, une situation qui ont été préalablement caractérisés et distingués des autres ; c'est ce qui fait dire à certains que la vision automatique est encore très loin de la vision humaine, capable de comprendre une scène complexe, ou un dessin humoristique, en une fraction de seconde, sans avoir été en quelque sorte « briefée » au préalable sur ce qu'elle est censée voir ou ne pas voir, c'est-à-dire ce dont elle doit simplement attester la présence ou l'absence. Pour le dire autrement, la vision artificielle, pour le moment du moins, fonctionne en univers clos, alors que la vision humaine fonctionne, du moins en apparence, en univers ouvert.

La vision artificielle a néanmoins accompli des progrès remarquables, et nourrit aujourd'hui des applications élaborées dans un grand nombre de domaines, de l'agriculture à la surveillance des routes, des bâtiments et des lieux publics, de la médecine à la fabrication industrielle ; alliée à la robotique, elle ne cesse d'étendre son domaine. Mesurées sur certaines échelles, dans certaines circonstances, les performances dans certaines tâches surpassent celles des humains. Mais elles rencontrent aujourd'hui des limites, qui sont de plusieurs ordres :

- les performances sont vulnérables à certaines petites variations, qui ne devraient faire aucune différence et qui conduisent à des résultats catastrophiques ; les réseaux de deep learning consacrés à l'identification d'images, par exemple, se laissent bernés par ce qu'on appelle les « images adverses » [*adversarial* en anglais, par référence au débat contradictoire en politique], variantes d'images standard que le réseau place dans une catégorie très éloignée de la catégorie correcte ;
- notre compréhension de ce qui fait que certaines architectures en réseau donnent de bons résultats reste très incomplète ; par delà les principes généraux, qui rendent intelligible la *possibilité* d'un apprentissage efficace, nous ne comprenons pas encore ce qui distingue, dans la classe des architectures, les bons élèves des élèves médiocres ;
- le lien avec la théorie des mécanismes visuels étant encore ténu, l'articulation des systèmes de vision artificielle avec des systèmes intégrant des connaissances ou des capacités inférentielles générales n'est généralement pas possible, ce qui empêche la vision artificielle, contrairement à la vision humaine, même chez le très jeune enfant, de bénéficier des ressources d'autres composantes, en particulier des connaissances de sens commun, et de créer d'elle-même des anticipations — en ce sens, la VA, si elle surpasse la vision naturelle dans certains cas, dans beaucoup d'autres ne s'en approche même pas.

4. La cognition réparée ou augmentée par le biais de son organe

Jusqu'à présent nous avons eu affaire à des systèmes artificiels de traitement de l'information qui interagissent avec l'humain ou le complètent « à cerveau constant » : ce sont, si l'on veut, des « prothèses » extérieures qui ne touchent pas à l'organe cognitif naturel. Nous allons maintenant parler de prothèses qui sont au contact de l'organe et qui le modifient ou le complètent par différents moyens, neuro-électriques ou

pharmacologiques. Parler de prothèse peut induire ici en erreur : s'il existe dès à présent des dispositifs, par lesquels nous commenceront, qui sont de véritables neuro-prothèses, la plupart des applications consistent en interventions, ponctuelles ou s'étendant dans le temps, visant à modifier le fonctionnement du cerveau, notamment en exploitant sa plasticité, c'est-à-dire sa capacité à modifier durablement sa structure (Bavelier et al. 2010, Hensch & Bilmoria, 2012)

4.1 Les interfaces cerveau-machine

Les interfaces cerveau-machine (ICM), désignés en anglais le plus couramment comme *brain-computer interfaces* (BCI), mais aussi *brain-machine interfaces* (BMI) ou encore *direct brain interfaces* (DBI), font l'objet d'un encadré dans le chapitre xxx. Nous nous bornerons donc ici à quelques remarques destinées à les situer dans un contexte plus large (v. Clerc et al. 2016, Brunner et al. 2015)

En premier lieu, les ICM se distinguent d'autres systèmes permettant d'actionner des membres en relayant les messages électriques jusqu'aux muscles : ils remontent à la source de l'action, à savoir l'*intention* formée par le sujet de faire bouger son bras, sa main, sa jambe..., en court-circuitant *complètement* la chaîne de commande neuromusculaire. Ils constituent pour cette raison une application et une confirmation beaucoup plus impressionnantes de la corrélation (certains diraient de l'identité) entre événement mental et processus cérébral : si l'intention de bouger peut causer le déplacement d'un membre, ou l'orientation du regard, alors l'intention est quelque chose de réel — comme le suppose bien l'introspection, mais comme l'atteste ici une expérience scientifique « en troisième personne », de manière objective. Mais c'est aussi ce qui leur donne leur intérêt pratique pour les patients qui ne sont capables d'aucune activité musculaire — comme dans le locked-in syndrome (rendu célèbre par le livre *Le Scaphandre et le papillon* de Jean-Dominique Baudy) et dans la phase terminale de la sclérose amyotrophique latérale.

En deuxième lieu, les ICM doivent n'être sensibles qu'aux intentions conscientes : s'ils me font lever le bras parce qu'inconsciemment (dans mon sommeil par exemple) je « veux » lever mon bras, ce ne sont pas des ICM ; or les ICM existent, donc l'intention consciente existe aussi, distincte de tout analogue inconscient concevable.

En troisième lieu, les ICM font partie d'une famille de dispositifs appelés *neuroprothèses*. Les ICM sont caractérisés par le fait qu'ils sont unidirectionnels : ils prennent certaines activités cérébrales en *input*, mais ne fournissent pas d'*output* au cerveau, contrairement aux dispositifs que nous allons considérer maintenant.

4.2 Les neuroprothèses sensorielles

Les neuroprothèses sensorielles pallient un organe sensoriel endommagé ou non fonctionnel. Ils résultent des efforts conjoints de l'ingénierie biomédicale, de la psychophysique et des sciences cognitives.

Les plus développées et diffusées parmi les neuroprothèses sensorielles sont les implants cochléaires (IC, en anglais CI). Ils permettent à un nombre croissant de sujets sourds ou malentendants de percevoir et comprendre le discours parlé sans avoir recours à la lecture sur les lèvres ou à la langue des signes. Apparus dans les années 1970, implantés à titre expérimental dans plusieurs centres européens, puis aux Etats-Unis, ils entrent dans la pratique clinique à partir de 1985 environ (Chouard 2010). Articulés à un système naturel extraordinairement performant (le système auditif), les IC sont un instrument complexe, comportant plusieurs composantes : la prise de son, le traitement du son, et la transmission physique du signal traité au nerf auditif par des électrodes implantées dans l'oreille interne (la cochlée). Chacune fait l'objet de perfectionnements constants faisant appel à des résultats technologiques mais également scientifiques de très haut niveau. Ce qu'un IC fournit au patient implanté est toutefois un signal très éloigné de la parole naturelle, appauvri considérablement sur le plan fréquentiel et temporel, mais paradoxalement intelligible dans le silence pour un sujet entendant entraîné. Cela s'explique par la grande redondance du signal de parole, mais également par le rôle-clé joué par des informations relativement grossières en reconnaissance de la parole. La situation est bien différente dans des conditions difficiles d'écoute, lorsque la parole est mélangée à du bruit ou distordue par de la réverbération. Les performances de l'IC diminuent très rapidement dans ce cas, révélant le rôle des informations fréquentielles et temporelles plus fines dans la reconnaissance robuste de la parole. Malgré les progrès considérables des quinze dernières années, l'IC n'est donc toujours pas une solution miracle. De nouvelles stratégies de codage - plus efficaces dans le bruit - restent à découvrir. Par ailleurs, l'IC exige dans tous les cas de l'utilisateur un effort d'apprentissage et un entraînement approprié. Cet effort n'est pas un aspect secondaire: pour que son expérience quotidienne soit satisfaisante et que la technologie soit adoptée, les contraintes d'acceptabilité doivent être prises en compte dès la phase de développement. On touche là à un principe de la recherche translationnelle, qui consiste à placer sur le même plan le contexte d'utilisation et la recherche de solutions technologiques. Un autre problème se pose, qui est celui de savoir s'il est *opportun*,

voire s'il est *légitime* d'implanter les sourds, particulièrement les enfants sourds. Nous reviendrons sur cette question dans la dernière partie du chapitre.

Les implants rétiniens (ou rétines artificielles — on parle aussi de systèmes de vision bionique) sont l'homologue des implants cochléaires dans la modalité visuelle. Les résultats encourageants obtenus en matière d'audition ont incité les chercheurs et les bio-ingénieurs à attaquer le problème de la déficience visuelle, ou plutôt à le réattaquer car certaines expériences avaient été faites dès les années 1930. Aujourd'hui, deux voies sont empruntées pour pallier certaines formes de cécité : les implants corticaux et les implants rétiniens. La première voie semble encore peu développée, compte tenu de la difficulté de faire exécuter par des dispositifs artificiels les traitements accomplis par les aires de haut niveau du cortex visuel. La seconde voie est plus prometteuse, précisément parce qu'elle enrôle les processus visuels naturels supérieurs ; elle consiste à « doubler » la rétine déficiente, soit en connectant la prothèse directement aux cellules ganglionnaires, soit en lui faisant jouer le rôle en amont des photorécepteurs. Comme dans le cas de IC, l'implant comporte trois grandes composantes : captation du flux lumineux, transformation en signal, transmission du signal au nerf optique selon l'une des deux configurations indiquées, épi-rétinienne ou sub-rétinienne. Les implants rétiniens ne sont entrés dans la phase clinique que récemment (les implantés se comptent par dizaines à ce jour, et non par centaines de milliers comme les IC), et sont beaucoup moins avancés que les implants cochléaires, pour des raisons que nous ne pouvons aborder ici, mais qui sont au moins en partie liées à la complexité computationnelle et neurale supérieure du système visuel.

4.3 De la thérapie électro-convulsive à la stimulation cérébrale profonde

La stimulation électrique du cerveau a une longue histoire, connue surtout sous la forme de thérapie électro-convulsive ou par électrochoc. Découverte par un psychiatre romain vers la fin des années 1930, cette thérapie ciblait les dépressions résistantes aux autres traitements (et prenait place aux côtés de deux autres formes de thérapie par choc chimique). Contestée, elle déclina au cours des années 1960, pour revenir lentement en faveur à partir du milieu des années 1970. Des progrès technologiques lui ont permis de reprendre sa place dans les années 1980 et certains perfectionnements technologiques en font une thérapie de choix, et de fait l'unique thérapie disponible pour un certain type de dépressions résistantes. On peut se demander toutefois, à juste titre, si l'on a là un exemple d'applications des sciences cognitives : ce n'est manifestement pas le cas, si l'on s'en tient à une définition stricte. Mais dans la mesure où la dépression a des manifestations cognitives — c'est même en ces manifestations qu'elle consiste, le fait qu'un certain type de choc électrique, comme une crise épileptique, peut faire disparaître pour un temps la dépression sévère, est potentiellement pertinent pour les sciences cognitives : le courant des idées passe, ou pourrait à terme passer, du terrain clinique aux sciences cognitives théoriques.

C'est aussi le cas d'une technique beaucoup plus récente, la stimulation cérébrale profonde (Sparing & Mottaghy 2008). Contrairement aux électrochocs, c'est une méthode invasive, nécessitant l'implant d'électrodes directement dans le cerveau. La stimulation cérébrale profonde a des effets sur la maladie de Parkinson, le tremblement, les troubles obsessionnels-compulsifs, la dystonie ; elle est aussi en phase de recherche dans le traitement de la dépression et de la douleur chronique, des troubles alimentaires, de la maladie d'Alzheimer et d'autres troubles d'origine neurologique. Il ne s'agit cependant pas de guérir la maladie, mais de réduire certains symptômes qui ne peuvent pas être contrôlés par les médicaments existants. Le contact entre ces recherches et les sciences cognitives, plus particulièrement les neurosciences, est plus étroit que précédemment, du fait que les électrodes pénètrent dans le cerveau, agissant de manière beaucoup plus spécifique que les électrochocs appliqués sur le scalp. Elles s'inscrivent dans le droit fil des recherches cliniques sur l'épilepsie, commencées par Wilder Penfield à Montréal dès les années 1940, poursuivies notamment à la Pitié-Salpêtrière dans les années 1960 par Jean Talairach et Jean Bancaud, qui mènent de la neurochirurgie aux neurosciences cognitives et les associent de manière toujours plus étroite aujourd'hui, dans le contexte des recherches sur la conscience et sur les états de conscience altérée ou diminuée.

4.4. L'entraînement cognitif

Les quinze dernières années ont vu se répandre l'idée que des formes d'apprentissage plus ou moins intensives, et plus ou moins directes, ont le pouvoir d'agir sur la structure et la fonction du cerveau adulte. Cette idée (parfois désignée comme *brain* — plus souvent que *cognitive* — *training*) se décline de deux façons principales : l'entraînement direct des fonctions exécutives et l'effet des jeux vidéo sur ces mêmes fonctions. Elle nourrit une véritable industrie qui vend des logiciels et des jeux à différents publics, plus particulièrement les enfants et les personnes âgées. Ces logiciels ciblent différentes capacités, telles que la mémoire à court terme, l'attention, la perception visuelle, voire l'intelligence fluide. Les jeux vidéo sont censés améliorer

l'attention, la coordination visuo-motrice ainsi que les fonctions exécutives.

Les résultats sont mitigés. Dans l'ensemble, c'est le scepticisme qui domine dans la communauté académique. Le jugement général sur lequel elle converge depuis quelques années est que l'entraînement améliore parfois nettement les performances dans les exercices sur lesquels les sujets sont entraînés, et dans des tâches proches, mais qu'il n'y a pas de transfert mesurable vers des fonctions plus générales. Dans le cas de l'entraînement à l'attention, des méta-analyses récentes indiquent des effets positifs, mais pas d'effet positif sur la mémoire de travail qui en est la cible. L'entraînement par les jeux vidéos semble produire des bénéfices pour les fonctions visuo-spatiales plus générales, mais là encore il faut interpréter les résultats avec prudence : leur confirmation par des laboratoires indépendants est nécessaire, et le lien *causal* entre l'utilisation de jeux vidéos et les fonctions exécutives ou visuo-spatiales reste à démontrer.

La situation actuelle est insatisfaisante : l'application des sciences cognitives est soit précise et fallacieuse (il n'est tout simplement pas vrai qu'elles suggèrent des technologies qui fonctionnent pour des raisons scientifiquement assignables), soit négative et générale (les bonnes pratiques méthodologiques mettent en cause l'efficacité prétendue du brain training). Reste toute une série de questions incontournables, telles que (1) L'idée même d'entraînement cognitif est-elle erronée, et si c'est le cas, pourquoi ? (2) Si ce n'est pas le cas, pourquoi les méthodes actuelles échouent-elles, et quel type de méthode serait-il susceptible de réussir ?

4.5 La stimulation cérébrale à la maison et les nootropiques chimiques

Deux autres formes d'intervention cognitive non invasive suscitent l'intérêt du grand public et soulèvent elles aussi des interrogations à la fois scientifiques et éthiques : les nootropiques et l'auto-stimulation cérébrale.

Les nootropiques ou nootropes sont une classe de médicaments agissant sur certaines fonctions cognitives, notamment la mémoire et l'attention. A l'origine utilisés dans les maladies d'Alzheimer et de Parkinson dans le but de réduire ou de retarder l'apparition de symptômes pathologiques, puis très (selon certains critiques, trop) largement administrés aux enfants présentant un trouble déficitaire de l'attention avec ou sans hyperactivité, leur utilisation s'est aujourd'hui répandue dans la population générale. L'espoir est alors d'améliorer les performances dans l'accomplissement de tâches qui sollicitent concentration, mémorisation, capacité d'assimilation, à un degré qui excède ce qui vient naturellement au sujet. Chez les lycéens, les étudiants et dans les professions intellectuelles, la prévalence de ces stimulants cognitifs ou « neuroenhancers », soit prescrits légalement, soit obtenus illégalement, est très loin d'être négligeable. Les estimations pour la population des Etats-Unis tournent autour de 12% d'utilisation permanente pour les 21-25 ans, et vont jusqu'à 25 à 35%, voire davantage, pour les étudiants de certaines régions et universités. C'est dire qu'on a affaire à un phénomène de société de première grandeur. Or on est loin de savoir mesurer les effets à court et à long terme des différentes classes de nootropes, leur innocuité et les mécanismes sous-jacents.

La stimulation magnétique transcranienne directe (tDCS) consiste en l'application de faibles courants électriques qui modifient de façon durable (mais non permanente) l'excitabilité des neurones corticaux. La tDCS a d'abord été utilisée comme outil d'exploration empirique en neurologie et dans le traitement des dépressions. Peu invasive et relativement économique, elle semble ne pas présenter d'effets secondaires majeurs (tels que des maux de tête ou le déclenchement de crises épileptiques). Pour ces raisons, elle a été utilisée dans une grande variété de contextes cliniques (Parkinson, Alzheimer, dépression majeure, schizophrénie, douleurs chroniques, AVC), mais aussi pour stimuler les fonctions cognitives chez le sujet sain (« neuro-augmentation »). Son utilisation est même devenue courante chez certains joueurs de jeux vidéo voulant améliorer leurs scores. Mais ici encore, l'opinion scientifique est très réservée sur son efficacité dans la plupart des contextes cliniques et en tant que stimulateur cognitif sur les sujets sains, ce qui pose le même genre de questions que les nootropes.

5. Contribution des sciences cognitives aux processus sociaux

Les sciences cognitives impactent la société de deux manières : par la voie scientifique, en interagissant avec les sciences sociales ; et par la voie pratique, en informant et en influençant certaines pratiques sociales. Les deux voies sont loin d'être étanches, d'autant que l'intrication entre sciences sociales et pratiques sociales, à notre époque et dans nos sociétés avancées, est très profonde. Il n'empêche que les échelles de temps sont différentes : lorsque les scientifiques interagissent entre eux, c'est dans le cadre de programmes de recherche à horizon assez lointain, tandis que lorsque la société fait appel aux sciences cognitives, c'est dans l'espoir de modifier à brève échéance certains comportements ou d'améliorer l'organisation sociale.

Le premier type d'interaction est abordé dans les chapitres « **Décision et rationalité** », « **De la coopération à la culture** ». Nous nous bornerons ici à exposer quelques exemples du second type, en présentant d'abord

trois applications ponctuelles réussies, puis quatre entreprises larges visant à enrichir et infléchir durablement tout un secteur de la vie sociale.

5.1 Applications ponctuelles

a. Les limites de l'attention et la conduite automobile

L'une des découvertes les plus solides des sciences cognitives porte sur l'attention : elle est très strictement limitée, ce qui fait que nous ne pouvons faire complètement attention à plusieurs choses à la fois, sauf si elles sont associées dans une longue pratique et se prêtent à une ségrégation des sous-tâches relatives à chaque activité (nous sommes capables de mener une conversation compliquée et de conduire sur une route encombrée sous une pluie battante, mais seulement si nous sommes bien entraînés à le faire). Le lecteur dira sans doute qu'il n'avait pas attendu que les sciences cognitives le lui apprennent : cela fait partie de notre psychologie naïve, et nous savons fort bien détourner l'attention de notre interlocuteur pour lui faire les poches ou l'empêcher de voir que nous reboutonnons une braguette oubliée. Qu'elle aille alors, notre lectrice, sur un des grands sites de vidéo et recherche « Selective attention test » ou « The invisible gorilla » (Simons & Chabris 2010) : elle comprendra alors, en regardant cette célèbre séquence, puis d'autres qui lui sont associées, combien notre psychologie naïve sous-estime le problème. La scène présentée illustre de manière spectaculaire ce qu'on appelle la cécité inattentionnelle : occupés à effectuer une tâche requérant une certaine concentration, nous pouvons ne pas percevoir une information visuelle pourtant parfaitement visible pour l'observateur non engagé. Un phénomène proche est celui de la cécité au changement, qui consiste à ne pas s'apercevoir d'un changement important dans une scène visuelle.

Le passage d'une intuition de sens commun à un fait scientifique parfaitement intelligible pour le profane est caractéristique d'un premier type d'application directe des sciences cognitives aux comportements quotidiens : chacun peut prendre conscience de la réalité de contraintes strictes, un peu comme lorsqu'on a obligé les candidats au permis de conduire à apprendre que pour immobiliser un véhicule se déplaçant à 90 km/h il faut une distance incompressible de 60 mètres — d'avoir des réflexes de Lucky Luke (et qui ne pense les avoir ?) n'y change rien. Une campagne publicitaire inspirée de la vidéo du « gorille invisible » a été produite en 2008 par la ville de Londres, et projetée dans les cinémas de la ville, dans le but d'alerter les conducteurs de voiture des risques de l'inattention. Ces risques étant considérablement accrus par l'utilisation du téléphone portable, la mise en évidence des limites strictes de l'attention humaine a incité le législateur à édicter une nouvelle réglementation. Dans un premier temps, on avait attribué les risques du portable au volant à l'occupation des mains du conducteur, et donc encouragé l'utilisation des écouteurs. Ce n'est que dans un deuxième temps, grâce à la diffusion de ces connaissances sur l'attention, que le législateur a compris que les accidents ne sont pas l'effet d'un manque de dextérité mais d'attention et qu'il faut interdire également les écouteurs. Remarquons que ces décisions ne découlent pas d'études statistiques permettant de corrélérer des accidents de voiture et l'utilisation du portable au volant, mais de la compréhension des mécanismes cognitifs impliqués.

b. La fiabilité de la mémoire et les témoignages oculaires

Peut-on croire les témoins oculaires ? Inutile de dire que nous avons de puissantes intuitions là-dessus, et de manière plus générale sur le fonctionnement de la mémoire. La preuve en est que les témoignages oculaires sont mobilisés dans une grande variété de situations judiciaires. La conviction intime d'avoir vu une voiture blanche s'arrêter devant la maison la nuit du crime semble garante de vérité, sauf cas très exceptionnels que les débats s'efforcent justement d'écarter. Cependant, l'expérience commune indique aussi que la mémoire peut nous induire en erreur, notamment lorsque la situation est complexe et/ou dramatique, lorsque nous avons un enjeu personnel, ou lorsque beaucoup de temps s'est écoulé.

Ici encore, ce que les recherches en psychologie cognitive ont établi, c'est la prévalence des souvenirs trompeurs (Loftus 2003). Dans une étude qui a fait date, des chercheurs ont montré à deux groupes de participants volontaires la vidéo d'un accident entre deux voitures. Au premier groupe, ils ont demandé à quelle vitesse circulaient les deux voitures quand elles se sont *touchées* ? A l'autre, à quelle vitesse elles circulaient au moment du *choc* ? On remarquera que les mots utilisés suggèrent des vitesses différentes avant l'accident. Une semaine plus tard, les chercheurs ont de nouveau convoqué les participants : ceux qui rapportaient du verre cassé sur la chaussée étaient plus nombreux dans le premier groupe que dans le second, alors qu'aucun éclat de verre ne figurait sur la vidéo. Il existe désormais des centaines d'études sur l'influence qu'une suggestion plus ou moins volontaire peut jouer sur les souvenirs. Un faux souvenir peut être induit non seulement par le langage utilisé, mais aussi par une fausse information glissée lors d'un interrogatoire. La leçon à retenir de cet ensemble de résultats est que notre mémoire, même dans des cas apparemment très simples,

n'est pas un enregistrement fidèle de la réalité. Cette leçon a eu des retentissements conséquents sur la manière dont la justice considère et prend en compte non seulement les témoignages oculaires mais aussi les souvenirs de victimes de crimes (Schachter & Loftus 2013). Les faux souvenirs et les faux témoignages ne sont pas nécessairement des mensonges, ils peuvent être le résultat du fonctionnement normal de notre mémoire (Costandi 2013).

c. Les nudges ou comment orienter les choix des gens sans leur faire violence

Sommes-nous disposés à donner nos organes si nous nous trouvons, suite à un accident, en état de mort cérébrale ? À cette question, les citoyens de certains pays répondent massivement « oui », ceux d'autres pays massivement « non ». On pourrait croire que les pays du oui sont typiquement des démocraties occidentales avancées, de préférence d'Europe du Nord et laïques, et que dans les pays du non ces traits sont moins nombreux, voire absents. Or ce n'est nullement le cas : aucun des critères culturels auxquels on pense naturellement ne permet de tracer la ligne séparatrice. Le critère opératoire est celui de la valeur par défaut : les pays dont la loi prévoit que le don d'organe est autorisé *sauf mention expresse du contraire* votent massivement « oui », et inversement lorsque la loi prévoit que le don d'organe n'est légal que s'il a été expressément autorisé par l'individu, le choix majoritaire est « non ».

Une fois de plus, le sens commun n'est pas pris totalement au dépourvu : en présence des faits, il construit une explication plausible. Les gens négligent généralement de modifier les valeurs par défaut, même s'ils sont vaguement conscients de l'existence de ces valeurs et de la possibilité de les changer, particulièrement lorsque l'éventualité que cela fasse une différence leur semble très peu probable : que nous nous trouvions un jour en état de mort cérébrale nous semble irréal, sauf peut-être si nous sortons d'un film d'Almodovar, ou travaillons dans un service de réanimation.

Ce qu'apportent les sciences cognitives, en l'espèce les travaux sur le raisonnement en situation d'incertitude engagés par Daniel Kahneman et Amos Tversky dès les années 1970 (Kahnemann, Slovic & Tversky 1982 ; Kahnemann 2011 ; voir le chapitre « Le nouveau paradigme du raisonnement »), c'est une compréhension détaillée de la manière dont la présentation d'un ensemble de possibilités influe sur nos choix. Ces travaux ont débouché sur la création d'un champ de recherches appliquées appelé, de manière trompeuse, neuroéconomie, « économie » parce qu'il concerne les choix d'agents en situation d'incertitude, « neuro » parce que les neurosciences apportent des ressources expérimentales nouvelles — il s'agit néanmoins, pour l'essentiel, de psychologie cognitive. Le versant appliqué de ces recherches concerne la manière dont ce qu'on appelle l'architecture de la décision conditionne, en probabilité, les choix que vont faire les agents. L'exemple de la donation d'organes en est une parfaite illustration. Un autre exemple courant est celui de la disposition des plats dans une cafétéria : en rendant plus facilement accessible les salades et les fruits, et plus difficilement les sandwiches au double bacon et les pâtisseries, on encourage les utilisateurs à manger « sain », sans les empêcher de manger « trop gras, trop salé, trop sucré » si tel est leur choix.

Ces questions s'inscrivent dans un effort mené depuis la naissance des sciences cognitives et de l'IA pour améliorer les décisions dans le domaine de l'action, de la politique à l'économie en passant par la santé (pour ce dernier cas, v. Gigerenzer & Muir 2011). Nous nous restreignons ici à un type d'intervention particulier.

Orienter la décision dans une certaine direction, sans empêcher l'agent de prendre une décision différente, ni lui imposer un coût sensible pour cela, c'est ce qu'on désigne à l'aide du terme anglais *nudge*, sans équivalent exact en français, qui signifie quelque chose comme « pousser doucement dans la bonne direction ». La stratégie des *nudges* fait désormais partie de la trousse à outils des pouvoirs publics, notamment au Royaume-Uni, où le gouvernement a créé en 2010 un organisme, désormais indépendant, la Behavioral Insights Team, plus connue comme la *Nudge Unit*, dont la fonction est de développer des dispositifs de *nudge*, dans toutes sortes de domaines, pour inciter un certain public à prendre des décisions jugées comme étant dans leur intérêt bien compris : par exemple, souscrire une assurance-vieillesse, payer ses impôts en temps utile, cesser de fumer, retrouver plus rapidement un emploi, vacciner ses enfants... (Thaler & Sunstein 2008). Les sciences cognitives apportent non seulement le cadre théorique et la notion de *nudge*, mais également les outils nécessaires à la conception (le design) de dispositifs adaptés au comportement visé et au public dont il s'agit de changer les habitudes. Nous évoquerons bientôt les questions éthiques posées par ce substitut « soft » de persuasion sociale.

5.2 Justice, commerce, politique, éducation

Comme on l'a expliqué au début du chapitre, il n'est aucune activité qui ne soit intéressée, potentiellement, par les avancées des sciences cognitives, mais c'est plus particulièrement le cas de celles qu'on pourrait appeler « cognitivement denses ». L'importance réelle, aujourd'hui, et les modalités pratiques

de leur apport varient beaucoup d'un champ à l'autre, comme le montrent les trois cas que nous allons présenter.

a. Le tournant cognitif de la justice

La justice peut-elle et doit-elle admettre la présentation de données cérébrales et cognitives devant une cour ? Telle est la question théorique générale qui se pose maintenant que de telles données peuvent être obtenues. Et, dans la mesure où la réponse est positive, comment encadrer l'utilisation de ces données, sur quels critères et à quel moment une cour de justice peut-elle décider de les prendre en compte et comment en particulier doit-elle les ordonner par rapport aux pièces et arguments traditionnels ?

La pertinence possible des données cérébrales repose sur l'hypothèse, défendue notamment par l'éminent neuroscientifique Michael Gazzaniga (Gazzaniga 2011), qu'il devient possible de lier certaines activités cérébrales à certains comportements, pensées ou états mentaux, en sorte que nos croyances actuelles sur le degré de contrôle qu'un prévenu a pu exercer sur ses actions pourraient être invalidées, et céder la place à des faits scientifiques, sur la base desquels on pourrait déterminer rigoureusement dans quelle mesure le prévenu devrait être tenu comme responsable.

Mais en attendant la réalisation éventuelle d'un tel programme, les images fonctionnelles du cerveau pourraient peut-être dès à présent entrer en jeu à d'autres niveaux d'un procès en justice. Premièrement, elles pourraient apporter la preuve d'une activité cérébrale anormale. Se pose alors le problème de déterminer ce qu'est une activité cérébrale normale, si toutes les catégories de personnes peuvent se référer à une seule norme, et la prévalence de certains types d'anomalies. Ces problèmes renvoient à la recherche fondamentale, encore à ses débuts, sur la caractérisation de cerveaux individuels, qui tranche avec la tradition en neuroimagerie qui consiste à étudier l'activité cérébrale en moyennant sur différents sujets de manière à faire ressortir l'universel humain. Si cette démarche fait sens pour découvrir les corrélats cérébraux des grandes fonctions cognitives partagées par tous les humains, elle est inappropriée pour révéler la manière dont une différence individuelle au niveau cérébral peut être liée à un comportement, pensée ou état mental spécifique. La recherche fondamentale est très loin de pouvoir fournir ce type d'informations.

Une seconde façon possible d'exploiter la neuroimagerie est pour la détection des mensonges, se substituant au traditionnel polygraphe abandonné à cause de son manque de fiabilité, ou en appui de cet instrument. Il est possible, grâce à des analyses de données complexes, de décoder dans certaines situations le contenu d'une activité cérébrale, de dire par exemple si quelqu'un pense ou se rappelle d'un accident de voiture plutôt que d'un coucher de soleil. Ce type d'analyse requiert une prise de données relativement longue et coûteuse, et des mesures répétées : il n'est pas possible de décoder le contenu d'une seule instance. Il est également impossible de déterminer si le contenu décodé est conforme à la réalité ou pas : le témoin pense peut-être à l'accident de voiture (comme il l'affirme), mais on ne sait pas si cette pensée indique qu'il a été témoin de l'accident. Là encore, les données d'imagerie cérébrale ne sont pas encore suffisamment fiables pour pouvoir être utilisées comme « détecteur de mensonges ».

Pourrait-on néanmoins faire valoir que les données issues des neurosciences ne sont pas moins sûres que les méthodes existantes pour identifier les coupables et les menteurs : pourquoi alors donc accepter les unes et refuser les autres ? La première objection est qu'elle place sur un pied d'égalité les deux types de données, alors que la saisie des facteurs en jeu, des motivations profondes des accusés, des circonstances exactes du crime sont peut-être plus sûrement cernés par les méthodes traditionnelles, assorties de tests de personnalité fiables. La seconde est qu'à partir du moment où on introduit les données neuroscientifiques, on risque de leur voir accorder a priori une supériorité sur les autres, en raison de l'aura qui entoure la science, et plus particulièrement les neuro-images en tant que produits de technologies de pointe. Les images en outre tendent à être perçues comme des données objectives alors qu'elles mettent en jeu tout un système d'interprétations.

b. Le tournant cognitif du commerce

Comme le postule la théorie du choix rationnel, le consommateur choisit en fonction de différents critères, tels que le prix et le gain attendu, mais contrairement à cette théorie, bien d'autres facteurs entrent en jeu, et bien d'autres objectifs que la maximisation de l'utilité espérée, en tout cas au sens classique. D'autre part, beaucoup de psychologues aujourd'hui estiment, avec Daniel Kahneman, que le système cognitif humain fait ses choix selon deux régimes, un régime lent (dit « système 1 ») de choix rationnel (au sens habituel), délibéré, algorithmique, d'un côté, de l'autre, par un régime rapide (dit « système 2 »), utilisant des heuristiques établies par l'évolution et adaptées à la plupart des situations communes, mais pas à toutes, notamment pas aux exceptions ni aux situations très éloignées des conditions primitives (voir le **chapitre** « Évolution et cognition »).

Ces processus sont désormais assez bien identifiés et permettent d'orienter un client potentiel vers une décision d'achat, notamment en manipulant son « système 1 » ; il devient également possible de prévoir, sur une base inductive, ses préférences, et de lui « recommander » tel et tel nouvel achat dans la ligne de son historique de commandes ; ou encore de tester, de manière plus scientifique qu'autrefois, les nouveaux produits qu'on songe à lancer sur le marché. Toutes ces applications relèvent de ce qu'on appelle souvent le « neuromarketing », de manière abusive dans la mesure où les neurosciences et la neuroimagerie ne constituent qu'une partie, relativement faible, de la trousse à outils de la spécialité. Il existe tout de même des applications des neurosciences. Un exemple connu est l'étude « Coke vs. Pepsi » où l'activité cérébrale des participants a été observée pendant qu'ils goûtent, en aveugle, les deux boissons. Pepsi semble activer plus puissamment des centres de la récompense ; mais lorsque les participants doivent évaluer le goût des deux boissons, non plus en aveugle, Coke remporte la compétition et l'activation cérébrale se modifie aussi par rapport à la phase en aveugle. Les auteurs de l'étude en concluent qu'un produit peut être préféré pour d'autres raisons que le goût, et que les justifications que nous donnons de nos choix peuvent différer de notre expérience « sub-personnelle » ou non consciente. D'autres études ont montré que non seulement nous pouvons être dans l'ignorance de ce qui nous fait choisir, mais que, contrairement à nos intuitions, nos choix peuvent créer nos préférences (en quoi nous serions logés à la même enseigne que d'autres primates).

c. Le tournant cognitif de la politique

Ce n'est certes pas d'aujourd'hui ni d'hier que date l'idée que la politique consiste en partie à prévoir puis à orienter les choix des citoyens, qui sont de ce point de vue dans la même relation aux hommes politiques que les consommateurs aux marchands. On ne s'étonnera donc pas d'apprendre qu'il existe un champ de recherches appelé « neuro-politique », mais à ce jour il est davantage l'affaire de recherches académiques, relevant des sciences politiques et de la psychologie sociale, que des équipes de campagne des partis et hommes politiques (Jost et al. 2014). Ceux-ci font néanmoins appel, de plus en plus, à des sociétés et experts de neuromarketing. Une fois encore, les méthodes employées, que ce soit dans le cadre académique ou dans le cadre directement politique, ne relèvent qu'en partie des neurosciences, mobilisant tout autant les méthodes de la psychologie appliquée.

La politique a de tout temps été affaire de manipulation : de la manipulation honnête parce qu'ouverte, reposant sur la persuasion rationnelle, jusqu'à la manipulation démagogique, voire insidieuse par activation d'affects à l'insu des citoyens. La neuropolitique contemporaine perfectionne ces méthodes en ciblant des micromarchés d'électeurs, à peu près de la même manière que le neuromarketing cible des micromarchés de consommateurs.

Pour se distinguer de ce genre de pratiques, les chercheurs intéressés par l'application des sciences cognitives à la politique désignent désormais leur champ de recherche comme *neurosciences politiques*. Les thèmes ne diffèrent pas ceux de la psychologie politique, un domaine beaucoup plus ancien : le racisme, les stéréotypes et les rapports entre groupes, les biais partisans et la manière dont les individus prennent connaissance de la situation politique est influencée par leurs intérêts et motivations, la nature des différences entre orientations ou tempéraments de droite et de gauche... Le tournant cognitif se manifeste à deux niveaux, d'ailleurs corrélés : la psychologie de la décision est profondément marquée par le programme « heuristiques et biais » de Kahneman et Tversky, tout particulièrement par la phase récente, celle de la théorie du double processus (*dual process*) qui donne aux processus automatiques une place essentielle ; et la neuroimagerie fonctionnelle apporte un nouveau type de données expérimentales, mais aussi des hypothèses sur l'architecture sous-jacente aux phénomènes identifiés par les psychologues.

Ces travaux constituent à la fois une application à la science politique et à la politique. Un exemple est celui de l'usage fait actuellement du « biais de négativité », qui fait que (selon les chercheurs qui l'étudient) nous sommes « câblés » pour être considérablement plus sensibles aux aspects, facteurs, événements négatifs que positifs (Baumeister & al. 2001). Pour la science politique et la science du management, il y a là une ressource explicative prometteuse pour l'étude du tempérament conservateur ou des dysfonctionnements de l'entreprise : en l'absence de contremesures drastiques, le biais de négativité oriente la pensée et les comportements dans une direction généralement néfaste. Mais le même genre de considération vaut pour la politique et le management ; c'est ainsi que Daniel Kahneman, dans une interview donnée en 2015, expliquait la large domination des « faucons » dans le conflit israélo-arabe, et préconisait des contremesures pour surmonter l'impasse actuelle.

d. Le tournant cognitif de l'éducation

On pourrait penser que l'éducation repose très directement sur la psychologie cognitive, et aujourd'hui, plus largement, sur les sciences cognitives, puisque son objectif est de conférer aux apprenants les meilleures

capacités cognitives possibles, ce qui compte comme une bonne capacité dépendant évidemment du contexte. Certes, une salle de classe n'est pas un laboratoire, un système éducatif n'est pas un institut de recherche : toutes sortes de dimensions (sociales, économiques, politiques, culturelles...) contribuent au phénomène éducatif *in situ*, il n'est personne pour l'ignorer. Cela n'ôte rien, semble-t-il, à l'idée que les sciences cognitives jouent un rôle central dans toute réflexion sur l'éducation : qu'elles sont, comme l'a proposé un chercheur du domaine, dans le même rapport à l'éducation que la biologie est à la médecine ; le psychologue Edward Thorndike parlait de dépendance : l'éducation *dépend* de la psychologie comme l'agriculture de la botanique et de la chimie. Ces parallèles, à la fois instructifs et optimistes, ne résolvent cependant pas la question : comme le disait le grand psychologue William James, s'adressant en 1899 à des professeurs de lycée, la psychologie est une science et la pédagogie un art : il n'existe donc pas de voie menant directement de la première à la seconde.

Plusieurs choses ont changé depuis James et Thorndike. La première est la preuve apportée par Jean Piaget, à partir des années 1920, que la psychologie scientifique peut apporter à l'éducation un éclairage essentiel, en se centrant sur le développement des capacités cognitives fondamentales du jeune enfant. Depuis Piaget, il est devenu impossible de mettre en doute le principe même d'une pertinence de la psychologie cognitive pour l'éducation, *via* les théories du développement et de l'apprentissage. La deuxième est la naissance des sciences de l'éducation, et en leur sein de courants plus expérimentaux qu'idéologiques, tels que la didactique des disciplines. Le troisième est l'émergence des sciences cognitives, qui donnent à la psychologie scientifique une impulsion considérable, lui apportant des ressources nouvelles, et lui permet de reprendre, en le modifiant, le programme de Piaget, et de systématiser les recherches empiriques des sciences de l'éducation. La quatrième, enfin, est que les dimensions sociales, politiques, culturelles de l'éducation ne sont plus considérées comme échappant au regard des sciences cognitives, même si celles-ci ne peuvent prétendre se substituer aux sciences sociales directement concernées.

Les conditions générales sont donc réunies pour des échanges fructueux entre sciences cognitives et éducation, échanges bilatéraux s'entend, car le monde de l'éducation (compris de manière large, en y incluant les formes non scolaires) a beaucoup à apporter à celui des laboratoires de sciences cognitives, et la recherche translationnelle est indispensable pour surmonter les multiples barrières qui se dressent entre les deux domaines. Un champ de recherche a bel et bien émergé, sous des intitulés tels que « Brain, Mind, Education », « neuroéducation » ou « neurosciences de l'éducation » (Bransford et al. 2000, Royal Society 2011, Pasquinelli 2014).

En pratique, les interactions se situent à différents niveaux, au sens exposé au §1.2 (Andler 2008). Les sciences cognitives ont pour premier effet de diriger l'attention vers les processus cognitifs à l'œuvre dans l'activité pédagogique, et sur la nécessité de compléter et de corriger les idées intuitives, généralement floues, que nous avons sur ce que sont ces processus et la manière dont ils conditionnent l'apprentissage chez les apprenants. Elles nous invitent aussi à nous interroger sur ce qui se passe dans la tête de l'enseignant. Nos intuitions ne sont pas seulement lacunaires ; elles sont souvent fausses : de le savoir, sans nécessairement savoir exactement en quoi, peut aider l'enseignant à accepter certaines difficultés, à se mettre en quête d'explications et à se défier des réponses toutes faites qu'internet et les ouvrages populaires lui suggèrent. À ce niveau de généralité, l'attitude expérimentale, la recherche de preuves, sont un garde-fou contre les fausses théories et contre les engouements pour des expériences ou des méthodes apparemment miraculeuses, mais qui ne doivent leur succès qu'à des facteurs tenus, volontairement ou non, dans l'ombre, et dont la validité ne s'étend pas au-delà des circonstances dans lesquelles elles ont été mises sur pied.

Un deuxième niveau d'implication porte sur les méthodes pédagogiques : les sciences cognitives mettent en évidence des contraintes précises, en matière d'acquisition des concepts, de mémorisation à long terme, d'attention, de charge cognitive, mais prennent également en compte la motivation, l'image de soi, le rôle de la métacognition et des théories spontanées de l'intelligence, etc. Ces contraintes ne dictent pas la conduite à suivre, les programmes, les méthodes, mais elles servent comme les cartes marines qui indiquent les écueils recouverts et suggèrent des itinéraires moins dangereux.

Au niveau supérieur se situent des applications directes de théories issues des sciences cognitives. Une première série d'exemples concerne l'apprentissage de la lecture, de l'écriture et de la numération. Nous en savons désormais assez, grâce à la psychologie et aux neurosciences cognitives, sur les mécanismes qui font partie des connaissances-noyaux, en place avant l'acquisition de ces aptitudes culturelles, impliquant le langage et les représentations symboliques, et aussi sur la perception visuelle et sur la perception auditive, pour recommander certaines méthodes et en décommander d'autres (par exemple, la méthode globale en matière de lecture). Cette ligne séparatrice est confirmée par des études empiriques rigoureuses. Tout cela reste bien entendu, comme toute production scientifique, sujet à révision, mais s'accompagne d'une garantie relative de validité : dans l'état présent de nos connaissances, et en l'absence de contre-arguments

convaincants et appuyés sur l'expérience, ce sont les prescriptions les plus fiables dont nous disposons. Elles s'appliquent à plusieurs populations : les enfants neurotypiques en situation normale ; en situation de décrochage ; les enfants « dys » atteints de dyslexie, dyscalculie, dyspraxie, dysgraphie...

Une deuxième série d'exemples porte sur les enfants autistes (selon la terminologie actuelle, atteints de trouble du syndrome autistique ou TSA). Ces enfants présentent un profil cognitif tout à fait particulier (qui n'est d'ailleurs pas exactement le même d'une sous-population à l'autre). Nos connaissances, quoique encore incertaines et incomplètes, permettent dans certains cas de construire des stratégies pédagogiques adaptées, et ouvrent en tout cas des pistes prometteuses.

Une troisième série d'exemples concerne les fonctions exécutives, la prise de décision, la planification, l'attitude devant le risque... Ce que nous savons sur ces mécanismes s'applique à différents profils d'apprenants, allant des enfants atteints de « learning disability » diffuse ou spécifique aux adolescents et aux individus atteints du syndrome déficitaire de l'attention et/ou d'hyperactivité (Diamond & Lee 2011). Ces travaux impliquent les neurosciences et la neuroimagerie, tout en relevant aussi de la psychologie cognitive. Mais les neurosciences commencent à intervenir directement, par exemple en précisant certaines contraintes physiologiques (sommeil, nutrition, rythmes scolaires...), en permettant le diagnostic précoce de certains déficits, ou en investiguant les avantages et les inconvénients du bilinguisme.

Enfin, les sciences cognitives apportent aux technologies de l'information et de la communication pour l'éducation (TICE) deux choses importantes :

- des principes pour la conception d'outils numériques, adaptés à la structure naturelle à laquelle ils doivent s'articuler, un peu comme une prothèse physiologique, s'agissant d'apprenants aussi bien typiques que déficients sensoriels, moteurs ou cognitifs ;
- une méthodologie d'évaluation rigoureuse, qui n'est pas propre aux sciences cognitives en tant que telles, mais qu'elles sont en pratique les mieux à même de déployer.

6. Les sciences cognitives dans le contexte social

Dans la section précédente, les sciences cognitives étaient considérées comme une source de connaissance pouvant alimenter et améliorer un certain nombre d'activités sociales organisées, depuis la conduite automobile jusqu'à l'enseignement. Il s'agissait, comme dans tout le chapitre, de questions de fait, parfois affectées d'incertitude. Dans cette dernière section, nous voulons évoquer des questions d'ordre normatif, c'est-à-dire engageant des jugements de valeur. Nous parlerons des *fausses* connaissances qui circulent sur le sujet de la cognition et du cerveau, et des conséquences éthiques, *bonnes* ou *mauvaises*, que peuvent avoir les connaissances, *vraies* et *fausses*, attribuées aux neurosciences cognitives plus particulièrement. Nous mentionnerons enfin, rapidement, les problèmes éthiques soulevés par certaines applications plus ou moins directes des sciences cognitives, conformément à ce que nous signalions d'entrée de jeu au §1.5.

6.1 Distortions et exagérations : les neurosciences cognitives source de fantasmes

L'intérêt, voire la fascination du grand public, des décideurs et des enseignants pour les sciences cognitives et plus particulièrement pour les neurosciences se sont développés au cours trente dernières années, et n'ont fait que croître pendant la « décennie du cerveau » (les années 1990). Cette « neurophilie » se nourrit des images colorées des aires actives du cerveau, qui ont supplanté l'atome de Bohr dans la représentation graphique de la science pour les foules (Roskies 2007). La prolifération des neuro-étiquettes pour les nouveaux domaines de recherche (neuro-éthique, neuro-éducation, neuro-marketing, etc.), la multiplication de projets, rapports et études sur les implications sociales, économiques, politiques et éducatives des neurosciences, tout cela témoigne de l'extension de ce mouvement à la sphère civile et politique.

a. Des méthodes pseudo-scientifiques

L'importance réelle des neurosciences cognitives ne soulève aucun doute. En revanche, elles donnent lieu à une exploitation sensationnaliste, ainsi qu'à des entreprises commerciales dénuées de fondement scientifique attesté (Robillard 2011). Un exemple en est Brain Gym®, une méthode d'apprentissage fondée sur l'idée que les différentes parties du cerveau doivent être « coordonnées » pour que l'apprentissage se fasse dans de bonnes conditions. Pour parvenir à cet état, il faut faire de courts exercices de gymnastiques stimulant de manière coordonnée les différentes parties du cerveau, boire de l'eau avant l'apprentissage, ou presser de fantomatiques « boutons » du cerveau pour faire remonter le sang vers l'organe de la pensée. Ces propositions sont en conflit avec des connaissances biologiques bien établies. Il n'y a par ailleurs aucune preuve que les

exercices proposés par cette méthode favorisent effectivement l'apprentissage (indépendamment du mécanisme de coordination supposé). Néanmoins, Brain Gym® et d'autres méthodes miracles sont globalement bien accueillies par le public et par le monde de l'éducation. Un second exemple est celui des méthodes VAK (visuel, auditif, kinesthésique) fondées sur l'idée, très à la mode, de « style d'apprentissage » ; le profil de chaque individu pourrait être quantifié par un questionnaire, à charge pour les enseignants d'adapter leurs pratiques éducatives au profil présumé de l'apprenant. Or on ne sait pas répondre à des questions pourtant essentielles, telles que celles-ci : La préférence révélée par le questionnaire s'applique-t-elle au-delà du questionnaire ? Une préférence pour une certaine modalité dans un domaine subsiste-t-elle dans un autre ? Les pratiques éducatives « adaptées » (en ce sens) donnent-elles de meilleurs résultats que d'autres pratiques ? S'il est rare que ces méthodes fassent l'objet d'études empiriques mesurant leur efficacité, les quelques méta-analyses de la littérature ne montrent pas d'effets positifs de ces méthodes sur l'apprentissage.

b. Les neuromythes

Un corollaire de la neurophilie est la prolifération de neuromythes (Pasquinelli 2014): des idées répandues dans le grand public, inspirées de résultats de sciences cognitives, mais déformées ou simplifiées au point d'être trompeuses. C'est ainsi que les travaux sur la spécialisation hémisphérique ont donné naissance au mythe que les gens sont plutôt « cerveau-droit » ou « cerveau-gauche », et que des exercices spécifiques permettraient d'atteindre un équilibre hémisphérique souhaitable. Les neuromythes peuvent également découler de théories scientifiques abandonnées en raison de l'émergence de nouveaux résultats, comme dans le cas de l'effet Mozart. Un article paru au début des années 1990 faisait état d'une légère augmentation du QI chez des adultes après écoute d'une sonate de Mozart, relativement à l'écoute d'une autre musique ou du silence. Suite à cette publication, des lois ont été promulguées dans certains territoires américains, imposant l'écoute de musique classique dans les écoles maternelles ou sa distribution aux nouvelles mères afin qu'elles la fassent écouter à leurs enfants. Or l'étude d'origine ne mentionnait aucunement les effets potentiels sur les enfants ni la possibilité que l'effet mesuré en laboratoire puisse donner lieu à des modifications à long terme de l'intelligence. De plus, les résultats de l'étude d'origine n'ont pu être reproduits, en sorte que l'existence d'un effet Mozart n'est plus admise dans la communauté scientifique. Cela n'empêche pas que des produits basés sur l'Effet Mozart (devenu une marque de fabrique) soient vendus à des millions d'exemplaires. Là encore, une idée séduisante par sa simplicité mais infirmée par la recherche s'est enracinée dans les croyances populaires.

Les neuromythes peuvent aussi se développer à partir de mauvaises interprétations des résultats expérimentaux. Par exemple, le mythe selon lequel nous utilisons seulement 10% de notre cerveau provient peut-être du croisement de considérations neuroanatomiques sur le rapport entre la quantité de cellules gliales et de neurones (ou entre substance blanche et substance grise), et de considérations mystiques ou parapsychologiques sur le potentiel inexploité de l'esprit humain.

c. Quels antidotes ?

Les cognitives partagent avec d'autres secteurs de pointe (notamment en informatique et IA ou en médecine) la rançon du succès. Elles ont le devoir de se prémunir autant que possible contre les falsifications et l'exploitation en appliquant les maximes générales de la bonne communication scientifique, et en encourageant les meilleures pratiques médiatiques : une bonne piste pour qui aspire à appliquer les sciences cognitives est de les diffuser de manière aussi sérieuse et efficace que possible, en formant le public et les médias ou en devenant soi-même médiateur.

Mais ce domaine a aussi des particularités, dues à la fois à sa proximité avec des valeurs essentielles (à cet égard il se situe sur le même plan que la médecine) et à ses compétences en matière de diffusion des croyances. Grâce notamment à l'apport de la pragmatique linguistique, par exemple, les sciences cognitives sont bien placées pour analyser certains mécanismes. Prenons un exemple. Dire qu'un processus mental particulier a lieu dans le cerveau est un truisme. Une image montrant qu'un certain processus mental se déroule à un certain endroit du cerveau n'apprend donc rien au profane. Ajouter des informations non pertinentes n'est pourtant pas sans conséquence : une importante littérature, notamment dans le domaine du marketing, montre que, selon les circonstances, cela peut soit renforcer soit affaiblir notablement l'effet des informations pertinentes. Lorsque l'information est enrobée du jargon des neurosciences, des explications circulaires sont prises à tort pour correctes. La simple présence d'une image du cerveau peut rendre une information plus persuasive.

Inversement, l'absence d'informations pertinentes, délibérée (au nom de la « lisibilité » pour le profane) ou non, peut déformer un résultat au point de conduire à une mécompréhension totale. Un exemple en est fourni par les images colorisées obtenues dans les comptes rendus d'expérience en neuroimagerie

fonctionnelle. L'image dépend du type d'analyse statistique et des conditions de contrôle adoptées, ainsi que du pouvoir de résolution (spatiale et/ou temporelle) du scanner utilisé (IRMf, TEP, MEG, EEG...). Il est donc faux qu'une neuroimagerie soit une sorte de photo du cerveau – au contraire, elle résulte de choix expérimentaux et analytiques. L'interprétation correcte d'une neuroimagerie implique nécessairement une connaissance de ces choix, qui ne sont quasiment jamais rapportés dans la presse, sans parler de l'incertitude scientifique qui entoure l'interprétation en question.

Enfin, l'opinion est particulièrement rétive au débriefing. Une fois assimilé un résultat scientifique annoncé dans les media, s'il se trouve démenti quelque temps plus tard, il est très difficile de le faire admettre par le public : les informations obsolètes restent gravées dans la mémoire collective. Une démonstration expérimentale de la résistance d'une idée à des informations contradictoires est fournie par une étude dans laquelle deux groupes de sujets lisent un article de presse au sujet d'un incendie dans un entrepôt. Le premier groupe reçoit une information selon laquelle les pompiers ont trouvé des matériaux inflammables, alors que le second groupe reçoit l'information selon laquelle ces matériaux, initialement pris comme inflammables, ne l'étaient en fait pas. On demande ensuite aux participants d'émettre une hypothèse sur la cause de l'incendie. Tous, y compris ceux qui ont été informés du démenti, proposent que la présence de matériaux inflammables est à l'origine du drame. Ici encore, les sciences cognitives sont à la fois victimes, comme d'autres disciplines, du phénomène, et équipées pour l'étudier et rechercher des antidotes.

On peut donc espérer que les sciences cognitives produiront à terme des antidotes à ces différents types de désinformation, en croisant les techniques générales déjà disponibles, notamment en matière de communication des résultats, avec une compréhension fine des mécanismes impliqués. Un beau problème d'application des sciences cognitives, avec retour des problématiques de terrain sur la recherche fondamentale.

6.2 Questions d'éthique

Les bonnes pratiques de communication scientifique font partie de l'éthique de la science. Les sciences cognitives sont comptables à titre égal aux autres sciences pour tout ce qui touche à la déontologie et plus largement à l'éthique de la science, dont on mesure de plus en plus l'importance.

Cela n'épuise pas le sujet. Il reste tout ce qui touche aux effets des sciences (cognitives en l'occurrence) sur les humains, pris individuellement et collectivement — ce n'est pas rien. Nous passerons rapidement en revue quelques exemples, choisis parmi les programmes mentionnés au cours du chapitre, puis concluons sur un ensemble de questions très particulières, qui constituent le champ de la neuroéthique.

a. Le principe de l'adaptation à l'utilisateur

Que les massicots soient ainsi faits qu'il faille appuyer simultanément sur deux boutons, un de chaque côté, en sorte d'éliminer le risque d'une main qui traîne quand la lame s'abat est évidemment une bonne chose. Que des interfaces soient conçus en sorte que la manipulation des instruments soit plus facile, plus « naturelle », est de bonne logique industrielle et commerciale. Pour les utilisateurs en situation de handicap, ils font toute la différence entre la dépendance et l'autonomie. Pour les pilotes d'avion et d'autres engins, pour les agents de centrales nucléaires ou de contrôle aérien, un affichage des paramètres mal adapté à la perception et à la cognition des utilisateurs peut conduire à une catastrophe, comme l'ont montré une longue série d'accidents. Il semblerait donc que tout ce qui se fait en matière d'« usabilité » dans le domaine des interfaces soit bon à prendre. C'est certainement l'opinion largement majoritaire, et on n'entend guère de voix dissidente.

Nous voudrions pourtant adresser au lecteur un défi, celui de réfléchir à ce que pourraient être les effets pervers d'une « suradaptation » des artefacts à l'agent humain. Voici quelques pistes — parmi d'autres — qu'il ou elle pourrait explorer :

- (i) L'idée d'une interaction naturelle renvoie à une notion de nature : pour le massicot par exemple, elle comprend la constitution corporelle (deux bras) et cognitive (le risque d'oublier un bras qui n'a pas de fonction dans la suite des mouvements requis). Mais dans d'autres cas, ce qui est pertinent dans notre nature n'a rien d'évident, et les concepteurs des interfaces n'en ont peut-être qu'une vue partielle et provinciale, par exemple, approximativement correcte pour des populations désignées par l'acronyme-valise « WEIRD » (*Western, Educated, Industrialized, Rich, Democratic*) introduit dans les débats méthodologiques sur l'expérimentation en psychologie cognitive et sociale, mais erronée pour d'autres utilisateurs potentiels.
- (ii) Rendre fluides, rapides et uniformes des processus qui prenaient jusqu'à récemment un certain

temps, engageaient des routines corporelles et cognitives spécifiques, pourrait avoir des effets négatifs sur le bénéfice tiré de ces processus. On sait par exemple que les étudiants américains qui prennent des notes sur leur ordinateur assimilent moins le cours que ceux qui prennent (plus lentement) des notes manuscrites. On peut se demander de même si les logiciels de dessin géométrique, permettant en quelques clics de réaliser une figure parfaite, sont aussi profitables aux élèves que la manipulation de la règle, de l'équerre et du compas. Certains écrivains reviennent à la plume, ou en sont restés à la vénérable machine à écrire, parce que les gestes nécessaires ont une texture particulière dont sont dépourvus les logiciels de traitement de texte.

- (iii) Meilleurs (au sens voulu par les concepteurs et les partisans de l'usabilité) sont les nouveaux artefacts de la vie quotidienne et de la vie professionnelle, plus nous en devenons dépendants, et par là dépendants des entreprises qui les fabriquent et les vendent. Compte tenu de leur taille gigantesque et de leur emprise mondiale, ces entreprises se voient ainsi détentrices d'importants leviers sur nos existences, et en mesure d'imposer leurs conditions, leurs normes, leurs valeurs.

b. Apprentissage automatique et données massives

L'AA soulève des interrogations proches de la piste (iii), et contrairement aux précédentes souvent discutées. Par construction, les grandes bases de données sont tributaires d'infrastructures colossales, consommatrice d'énergie à grande échelle, appelant de coûteuses mesures d'entretien et de sécurité. Ces infrastructures sont contrôlées par des entreprises globales plus puissantes que nombre d'États, et dans une certaine mesure par des États. Les données massives devenant le passage obligé pour tout un ensemble d'activités d'intérêt collectif, ces activités se trouvent à leur tour sous le contrôle de ces entreprises et de ces États, échappant ainsi au moins en partie à la société civile.

Une première conséquence est qu'en matière de données personnelles (et une très forte proportion des données collectées le sont), la protection de la vie privée ne peut être que très imparfaite. Une deuxième, bien connue, est que ces données sont une source de richesse considérable sur laquelle les entreprises exercent un droit de ponction illimité. Une troisième est que dans l'état présent du droit et des usages, et vu la situation actuelle d'oligopole des infrastructures, il est impossible de tirer des données existantes sur internet, par croisement, des informations qui seraient socialement utiles. Un exemple proposé dans un état de l'art récent est celui de l'identification précoce et de la maîtrise de pandémies, dont on connaît le risque accru aujourd'hui. Il serait techniquement possible, et même assez facile, de croiser les informations de localisation provenant de sources telles que les téléphones portables, les transactions par carte de crédit, les caméras de surveillance, avec les informations sanitaires (services d'urgence des hôpitaux, dossiers médicaux) pour identifier les foyers d'épidémie et prévenir la contagion en contactant en temps réel les individus exposés. Mais c'est impraticable aujourd'hui, parce que les questions de confidentialité ne sont pas maîtrisées et plus encore parce que les données relèvent de régimes de propriété hétérogènes, les détenteurs légaux n'ayant aucun intérêt à unir leurs efforts, et souhaitant au contraire s'entourer d'une clôture aussi hermétique que possible pour éviter le piratage ou encore la prise de contrôle par des instances politiques ou policières.

Une autre question est celle-ci : les données massives, traitées par AA, font surgir des informations sur les individus et sur les groupes sociaux qui sont inconnues d'eux. C'est l'image en miroir du problème de la confidentialité. Dans un cas, ce que je sais de moi-même et qui ne regarde que moi devrait être inaccessible à toute personne ou instance non autorisée, mais est en réalité connu d'une instance tierce ou risque de l'être ; dans l'autre, ce que sait de moi une instance tierce, je devrais en avoir connaissance (et savoir que quelqu'un d'autre le sait), alors que je l'ignore — et de même, plus gravement peut-être encore, pour un groupe humain.

c. Implants cochléaires et culture sourde

Faut-il, chaque fois que c'est possible, munir un individu sourd d'implants cochléaires ? Alors qu'on pourrait penser que la question ne se pose pas, puisque l'individu implanté gagne quelque chose d'incalculable, l'accès à la parole d'autrui et l'accès en tant qu'égal au monde des entendants, certaines communautés de sourds de naissance le contestent, au nom de l'appartenance à une culture particulière. La dotation précoce des implants cochléaires (pour des enfants à l'âge de 2 ans, par exemple) crée selon eux une forme inédite de stigmatisation. En grandissant, sourd, l'enfant s'intègre à une culture particulière, la culture sourde, ayant son langage et ses systèmes de communication. Les implants lui imposent, avant qu'il n'ait l'âge pour choisir, l'appartenance à une autre culture, celle du langage oral et de l'écrasante majorité des non-sourds, avec ses valeurs et ses préventions à l'égard du « handicap », dans laquelle il ne peut être justement que, toujours et encore, « handicapé », dysfonctionnel (vu les limites techniques des implants) : il quitte une culture sans pouvoir se fondre dans une autre (Sparrow 2005). La « culture des implants » serait donc incapable de

comprendre l'expérience de la surdit  autrement que comme une maladie   soigner, et elle imposerait une vision univoque de la normalit . En outre, les implants cr eraient un d calage entre l'enfant et ses parents sourds-n s. Loin donc d'aller de soi, l'implantation soul ve un ensemble de probl mes  thiques : celui du libre choix, notamment dans le cas d'enfants ; celui du consensus inform , concernant des parents qui peuvent ne pas disposer de tous les  l ments pour le choix ; et enfin celui de l' quilibre entre co ts et b n fices – vu que les implants, bien qu'efficaces, sont encore incapables de produire une exp rience sensorielle pleinement satisfaisante. A cela s'ajoute le probl me de l'acc s   ce genre de technologie. Et comme c'est souvent le cas en mati re biom dicale, les d cisions ne peuvent pas attendre  ternellement : il existe une p riode critique pour le d veloppement du syst me auditif au niveau cortical, ce qui impose une forte pr f rence pour une implantation avant 3 ans et demi, afin de r duire les risques li s   la privation sensorielle.

d. Les nudges et le paternalisme libertarien

Est-il licite de faire le bien des gens   leur insu ? N'est-ce pas l  une forme de paternalisme, d'autant moins acceptable qu'il se dissimule, contrairement aux pratiques traditionnelles de la persuasion sociale ? « Quand les parents boivent, les enfants trinquent »  tait le th me d'une campagne antialcoolique de 1957, appuy e par une affiche inoubliable. Une campagne de ce genre ne fait pas myst re de son intention, au contraire : son message doit  tre totalement explicite – en l'esp ce, il  quivaut   peu pr s   « L' tat vous recommande de ne pas consommer (trop) d'alcool ». Le nudge se situe   l'extr me oppos  : il est silencieux, et son efficacit  tient en bonne partie   son silence. Il faut un effort pour se rendre compte, premi rement que les p tisseries sont moins faciles d'acc s que les salades de fruit, et deuxi mement que ce n'est pas par hasard, que cette disposition pousse vers une alimentation jug e plus saine par quelqu'un dans les sph res sup rieures.

Les partisans des nudges font d'abord valoir que, contrairement au paternalisme traditionnel, qui est fonci rement autoritaire m me lorsqu'il se dissimule derri re une certaine bonhomie, le paternalisme du nudge pr serve l'autonomie de l'agent : rien n'emp che de choisir la p tisserie, ni de d cocher la case « Oui, je souscris   un plan-retraite » ; de fait, l'effort requis est minimal. En ce sens, si paternalisme il y a, il est d'une nature particuli re, il se qualifie comme « libertarien » (un terme d    Cass Sunstein et Richard Thaler). Une seconde riposte est que certains choix sont objectivement pr f rables, et que le devoir de la soci t  est d'aider   les faire ceux qui pour une raison ou une autre n'en sont pas capables. Ils ajoutent parfois que ce devoir s'adresse non seulement aux individus concern s, mais   la soci t  toute enti re : ne pas « nudger » les fumeurs   moins fumer, par exemple en disposant les cigarettes dans un coin mal  clair  tout au fond du magasin, c'est renoncer   une baisse statistique du nombre de cancers du poumon, donc inversement imposer une d pense  vitable   la soci t  enti re. S'il faut pour cela forcer un peu la main des fumeurs, f t-ce insidieusement, c'est l  le co t   payer.

Mais aux yeux de ses critiques, ce paternalisme insidieux est inadmissible. Il n'est pas possible, selon eux, de d terminer l'int r t sup rieur d'autrui de mani re objective. Et certains placent le respect scrupuleux de la libert  individuelle au-dessus de toute autre consid ration : aucune manipulation, si minime et non contraignante qu'elle puisse  tre, n'est acceptable, du fait m me qu'elle proc de d'une *intention* de restreindre cette libert .

e. L'augmentation cognitive

Les techniques d'augmentation cognitive (*cognitive enhancement*) de sujets sains, par voie m dicamenteuse, par simulation magn tique transcranienne directe ou par tout autre mode, existant ou   venir, d'intervention sur le syst me nerveux central, posent toute une s rie de probl mes d licats.

Certains de ces probl mes naissent de la difficult    tracer des fronti res nettes. Ainsi, ces techniques sont g n ralement consid r es comme acceptables lorsqu'elles ont une vis e th rapeutique : si elles permettent   un patient que son affection handicape, par exemple sur le plan scolaire ou professionnel, elles semblent non moins l gitimes que d'autres formes de traitement. S'agissant de sujets sains, ce n'a pas le m me degr  d' vidence. Mais entre le pathologique et le normal existe une zone grise, plus  tendue encore que dans la m decine du corps.

Une autre fronti re probl matique concerne le caract re direct ou indirect de l'intervention. S'entra ner   conna tre les dizaines de milliers de voies de Londres (comme y  taient oblig s les chauffeurs de taxi londoniens avant le GPS) entra ne une augmentation du volume de l'hippocampe, et personne ne songe   contester la l gitimit  de cette modification. S'il devenait possible d'obtenir une semblable augmentation par exposition   la tDCS, on pourrait se demander au nom de quoi on devrait s'en abstenir ; mais aussi, o  se situerait une technique (encore imaginaire) reposant sur des exercices de respiration, sur une immersion dans

de l'eau glacée, ou sur un régime alimentaire à base de gingembre ?

Une troisième distinction porte sur la valeur intrinsèque et la valeur instrumentale de l'augmentation. Qu'un individu choisisse librement de recourir à une technique d'augmentation dans le seul but de jouir d'une cognition plus performante (être ou se sentir plus créatif, capable d'une meilleure concentration, d'avoir une meilleure mémoire...) ne soulève pas d'objection : la poursuite de cette valeur intrinsèque n'est pas moins légitime que de s'exercer pour renforcer ses abdominaux. En revanche, si l'intention est d'acquérir un avantage sur des concurrents qui n'ont pas accès à la technique ou ne peuvent ou ne souhaitent pas y avoir recours, la pratique devient éthiquement problématique, mais pas nécessairement condamnable : les méthodes classiques d'éducation sont dans le même cas, compte tenu de l'accès inégal des catégories socio-professionnelles et des communautés d'appartenance. On songe au débat sur le dopage sportif : s'entraîner est normal, se doper est tricher. Est-ce si évident que cela ? Et l'augmentation cognitive relève-t-elle de la même logique ?

Un autre ensemble de problèmes se posent : ce sont ceux dont s'occupe l'éthique médicale. Ils portent sur les risques encourus et sur l'équilibre entre coûts et bénéfices à court et long terme, sur l'information des personnes concernées, sur la légitimité d'une décision pour autrui (cas des enfants voire de très jeunes gens enrôlés dans des équipes nationales préparant des compétitions de mathématiques ou d'échecs), etc. Il est devenu clair en éthique médicale que ces questions ne peuvent être résolues une fois pour toutes et pour tous les cas : elles nécessitent un débat, pour chaque type de situation, en étroite contact avec les pratiques. C'est pourquoi l'augmentation cognitive, qui ne relève d'ailleurs pas directement de l'éthique médicale dans la mesure où, comme le dopage sportif, elle concerne aussi des sujets sains, donne lieu à une réflexion spécifique.

f. La neuroéthique

Ce terme désigne un domaine désormais assez développé, dont les contours varient selon le contexte. Au sens large, il inclut toutes les questions d'ordre éthique soulevées par les sciences cognitives (le préfixe « neuro » jouant ici aussi le rôle de pointeur vers le champ dans son ensemble), leur réception, leur retentissements sociaux, leurs applications... En ce sens, toute la présente sous-partie (6.2), et même la partie entière (§ 6) auraient pu s'intituler « Neuroéthique ».

En un sens plus étroit, la neuroéthique concerne les applications des neurosciences à tous les domaines : elle inclut par exemple la question de l'augmentation abordée dans la section précédente, mais également toutes les interventions médicales sur le système nerveux, de la thérapie électro-convulsive à la neurochirurgie ou aux traitements médicamenteux en psychiatrie, le diagnostic, le traitement et l'usage expérimental des patients atteints de maladies neuro-dégénératives, les syndromes spécifiques tels que l'autisme, Down, Williams, etc.

En un troisième sens, la neuroéthique concerne la portée éthique et anthropologique du programme descriptif et explicatif des neurosciences cognitives : ici le préfixe « neuro » doit être pris au sérieux, même si la problématique concerne l'ensemble des sciences cognitives et la philosophie morale, parce qu'elle naît du projet même de faire dire à l'étude scientifique du *cerveau* quelque chose d'important sur l'être humain. C'est ce troisième sens qui va nous occuper dans les lignes qui suivent, et nous distinguerons trois dimensions de ce projet (v. notamment Farah 2010).

La première correspond à l'idée que les neurosciences pourraient rendre compte, ou du moins contribuer à éclairer ce phénomène à la fois central et mystérieux de l'existence humaine qu'est la morale (terme qui est pris ici comme synonyme de l'éthique, quoique dans d'autres contextes les philosophes les distinguent nettement). Dit brutalement, en ce sens la neuroéthique consiste à développer une théorie neuroscientifique de la morale, de la même manière que les neurosciences développent des théories du mouvement volontaire, de la décision ou de la perception visuelle (Gazzaniga 2005).

Une telle théorie, si elle existe, a pour effet de « naturaliser » la morale, de la situer dans la sphère des processus naturels. Comme on l'a vu dans le chapitre « De la coopération à la culture », les ressources de la biologie évolutive sont également enrôlées dans une entreprise de cet ordre. Se pose alors la question de savoir si ces entreprises sont en voie de réussir. Si ce n'est pas le cas, quelles en seraient les raisons ? L'idée même d'une naturalisation de la morale est peut-être contradictoire, ou bien les neurosciences, la biologie évolutive et la philosophie n'ont pas encore atteint un degré suffisant de maturité. Si, à l'inverse, la réussite semble proche, voire déjà acquise, quelles en sont les conséquences anthropologiques ? L'image de l'humain que nous ont léguée les traditions philosophiques et théologiques doit-elle subir une révision radicale ? Tout particulièrement, les notions communes de responsabilité et de liberté (ou de libre arbitre) doivent-elles être considérées comme caduques ? Concrètement, par exemple, un meurtrier pourra-t-il dire au tribunal que c'est son cerveau qui a commandé son geste criminel, et qu'il faut donc soigner son cerveau plutôt que trancher la tête (ou enfermer le corps) qui le contient ? Un sauveteur héroïque devra-t-il se contenter d'un rapport

d'activité, l'idée même de mérite et de récompense s'étant révélées dénuées de sens ? Cet ensemble de questions est principalement l'affaire des philosophes, mais requiert une compréhension approfondie des méthodes et des résultats des neuroscientifiques. Il constitue la deuxième dimension du projet général.

La troisième concerne les échos de cette problématique dans la conscience commune. À partir du moment où se répandent dans la population générale les hypothèses d'un déterminisme cérébral (renforcées par celles d'un déterminisme génétique, causalement antérieur), la conscience que des philosophes éminents y souscrivent et pensent en effet que nous ne sommes responsables ni de nos fautes ni de nos bonnes actions, l'organisation sociale actuelle, fondée sur l'idée de responsabilité et de mérite, ne risque-t-elle pas de s'écrouler ? Que faire de l'instruction civique et morale des enfants ? Quels principes doivent-ils désormais guider la politique étrangère des États ? Ou bien alors la bonne question n'est-elle pas de savoir comment faire comprendre que les avancées des neurosciences ne modifient pas *fondamentalement* les intuitions communes qui orientent, par delà la diversité des codes moraux, notre conduite, tout en leur apportant un éclairage nouveau et en modifiant à *la marge* un certain nombre de jugements ?

Avec ces dernières remarques, nous sommes revenus à des questions théoriques, qui peuvent sembler bien éloignées des projets technologiques, sociaux ou politiques que les sciences cognitives permettent ou promettent de développer. En réalité, de même que les applications d'une science quelconque, et tout particulièrement des sciences cognitives, exigent des aller-retour que doit systématiser la recherche translationnelle, comme nous l'avons vu, de même les questions générales posées par l'application des sciences renvoient à des questions générales que traitent les sciences théoriques et la philosophie. Cette intrication du « pur » et de l'« appliqué » est désormais reconnue, et dans notre domaine plus que dans tout autre elle est un moteur puissant qu'il ne faut surtout pas chercher à brider.

Bibliographie

Liste A : Pour aller plus loin

- Chabris, Christopher F. & Simons, Daniel J. (2010). *The invisible gorilla*. New York, NY: Crown.
- Clerc, Maureen et al. (eds.) (2016), *Brain Computer Interfaces*, 2 vol., New York : Wiley
- Farah, Martha J., (ed.) (2010). *Neuroethics: An Introduction with Readings*. Cambridge, MA : MIT Press.
- Frith, Uta & Blackmore, S-J. (2005). *The Learning Brain*. London: Wiley-Blackwell.
- Gazzaniga, Michael S. (2005). *The Ethical Brain*. New York : The Dana Press
- Kahnemann, Daniel (2011). *Thinking, fast and slow*. New York : Farrar, Straus and Giroux.
- Norman, Don (2013). *The Design of Everyday Things* (Revised and expanded ed.). New York : Basic Books
- Pasquinelli, Elena (2014), *Du labo à l'école : science et apprentissage*, Paris : Le Pommier
- Simon, Herbert A. (1969/1996). *The Sciences of the Artificial*, 3^e éd. refondue, Cambridge, MA : MIT Press ; trad. fr. de la 3^e édition : *Les sciences de l'artificiel*, Paris : Gallimard, coll. Folio, 2004
- Thaler, Richard H., Sunstein, Cass R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press

Liste B : Toutes les références du chapitre, y compris Liste A, à inclure dans la biblio générale

- Andler, Daniel (2008). Sciences cognitives et éducation: une relation sérieuse. In D. Andler & B. Guerry (Eds.): *Apprendre demain*. Paris : Hatier. Disponible sur son site.
- Ariely, Dan (2008). *Predictably Irrational*. New York : Harper Collins.
- Baumeister, Roy F. & al. (2001), Bad is stronger than good, *Review of General Psychology* 5, 4, 323-370.
- Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: From molecular to behavioral interventions. *Journal of Neuroscience*, 30(45), 14964–14971.
- Bransford, John D. et al. (eds.), (2000). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press, 2^e éd. ; disponible à <http://www.nap.edu>
- Brunner, C., et al. (2015). « BNCI Horizon 2020: towards a roadmap for the BCI community », *BCI Journal*, 1, 3-4, 1-10.
- Burdea, Grigore C. & Coffet, Philippe (2003). *Virtual Reality Technology*, Wiley-IEEE Press, 2^e éd.
- Chouard, C.-H. (2010) . Histoire de l'implant cochléaire. *Annales françaises d'oto-rhino-laryngologie et de*

pathologie cervico-faciale 127, 288—296.

Clerc, Maureen et al. (eds.) (2016), *Brain Computer Interfaces*, 2 vol., New York : Wiley

Costandi, M. (2013). Evidence-based justice: corrupted memory. *Nature*, 500(7462):268-70.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old, *Science*, 333(6045), 959–964.

Eichenbaum, Adam, Bavelier, Daphne, & Green, C.Shawn (2014). Video Games: Play That Can Do Serious Good , *American Journal of Play*, 7(1), 50-72

Farah, M.J., (Ed.) (2010). *Neuroethics: An Introduction with Readings*. Cambridge: MIT Press.

Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms that make Sense of Data*, Cambridge : Cambridge University Press.

Gazzaniga, M. S. (2005). *The Ethical Brain*. The Dana Press

Gazzaniga, M. (2011). Neuroscience in the courtroom. *Scientific American*, 304, 54 – 59.

Gibson, J.J. (1977), The Theory of Affordances. In R. Shaw and J. Bransford (eds.) *Perceiving, Acting, and Knowing*. Lawrence Erlbaum Associates.

Gigerenzer, G., & Muir Gray, J. A. (Eds.). (2011). Better doctors, better patients, better decisions: Envisioning health care 2020. Cambridge, MA: MIT Press.

Goodfellow, Ian, Bengio, Yoshua & Courville, Aaron (2016), *Deep Learning*, Cambridge, MA : MIT Press

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3, 446-474.

Hensch, T. K., & Bilimoria, P. M. (2012). Re-opening Windows: Manipulating Critical Periods for Brain Development. *Cerebrum: The Dana Forum on Brain Science*, 2012, 11.

James, W. (1925). Talks to the teachers on psychology: and to students on some of life's ideas. New York: Henri Holt and Company

Jost, John T., Nam, H. Hannah, Amodio, David M., Van Bavel, Jay J.(2014), Political neuroscience : The beginning of a beautiful friendship, *Advances in Political Psychology*, Vol. 35, Suppl. 1,

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kahnemann, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Loftus, E. (2003). Our changeable memories: legal and practical implications. *Nature Reviews Neuroscience*, 4, 231-234.

Norman, Don (2013). *The Design of Everyday Things* (Revised and expanded ed.). New York : Basic Books

Pasquinelli, Elena (2014), *Du labo à l'école : science et apprentissage*, Paris : Le Pommier

Pasquinelli, E. (2015), *Mon cerveau, ce héros*, Paris: Le Pommier.

Peter, C. & Beale, R. (Eds.) (2008). *Affect and Emotion in Human-Computer Interaction. From Theory to Applications*. Springer.

Robillard, J.M., Illes, J. (2011). Lost in translation: neuroscience and the public. *Nature Reviews Neuroscience*, 12, 118.

Roskies, A.L. (2007). Are neuroimages like photographs of the brain?, *Philosophy of Science*, 74: 860-872.

Royal Society (2011). Neuroscience: Implications for education and life-long learning. London: Royal Society.

Russell, Stuart & Norvig, Peter (2009), *Artificial Intelligence, A modern approach*, 3e ed., Pearsons,

Schacter, Daniel L, & Loftus, Elizabeth F. (2013) Memory and Law: What Can Cognitive Neuroscience Contribute? *Nature Neuroscience* 16(2): 119–123.

Simon, Herbert A. (1969/1996). *The Sciences of the Artificial*, 3^e éd. refondue, Cambridge, MA : MIT Press ; trad. fr. de la 3^e édition : *Les sciences de l'artificiel*, Paris : Gallimard, coll. Folio, 2004

Simons, D. J., & Chabris, C. F. (2010). *The invisible gorilla*. New York, NY: Crown.

Sparing, R., Mottaghy, F.M. (2008). Noninvasive brain stimulation with transcranial magnetic or direct current stimulation (TMS/tDCS)—From insights into human memory to therapy of its dysfunction. *Methods* 44 (4): 329–37.

Sparrow, R. (2005). Defending Deaf Culture: The Case of Cochlear Implants. *Journal of Political Philosophy*, 13, 2, 135–152.

Sun GD, Wu YC, Liang RH et al. (2013) A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology* 28(5): 852-867 Sept. 2013.

Thaler, Richard H., Sunstein, Cass R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press

Tversky, B. (2011). Visualizations of thought. *Topics in Cognitive Science*, 3, 499-535.