

L'esprit étendu

Andy Clark et David J. Chalmers

1. Introduction

Où s'arrête l'esprit et où commence le reste du monde?

Cette question appelle deux réponses élémentaires. Certains acceptent la frontière de la peau et du crâne et affirment que ce qui se trouve en dehors du corps est également en dehors de l'esprit. D'autres sont convaincus par les arguments consistant à dire que le sens des mots que nous utilisons ne se trouve pas « seulement dans notre tête » et tiennent pour vrai que cet externalisme du sens mène à un externalisme de l'esprit. Nous proposons une troisième position et défendons un externalisme tout à fait différent : *un externalisme actif*, fondé sur le rôle actif de l'environnement dans la conduite des processus cognitifs.

2. La cognition étendue

Prenons trois exemples de résolution de problème :

1) Une personne est face à un écran d'ordinateur sur lequel s'affichent des images de formes géométriques en deux dimensions. Elle doit répondre à des questions sur la possibilité de faire entrer ces formes dans des espaces correspondants. Pour vérifier ses hypothèses la personne doit faire tourner les formes dans sa tête pour les aligner avec les espaces.

2) Une personne est face à un écran d'ordinateur identique au précédent, mais cette fois-ci elle a le choix entre faire tourner physiquement l'image à l'écran en appuyant sur un bouton ou de faire tourner l'image dans sa tête comme dans le cas précédent. On peut faire l'hypothèse, réaliste, que l'opération de rotation physique suppose un gain de vitesse.

3) Dans un futur cyberpunk, un agent est face à un écran d'ordinateur similaire. Mais cet agent possède un implant neuronal qui lui permet d'effectuer la rotation aussi rapidement que l'ordinateur dans l'exemple précédent. L'agent doit encore choisir quelle ressource interne utiliser (l'implant ou la bonne vieille rotation mentale), sachant que chaque ressource suppose des exigences différentes en termes d'attention et de concurrence avec d'autres activités mentales.

Quelle part de *cognition* est à l'œuvre dans chaque cas ?

Nous pensons que les trois cas sont similaires. L'exemple numéro 3, celui de l'implant, semble être sur le même plan que le cas numéro 1. Et le cas numéro 2, avec le bouton, suppose la même structure informatique que le cas numéro 3, bien qu'elle soit distribuée entre l'agent et l'ordinateur plutôt qu'interne à l'agent. Si la rotation est cognitive dans l'exemple 3, pourquoi considère-t-on que le cas numéro 2 est fondamentalement différent ? On ne peut se contenter de citer la frontière peau/crâne, puisque c'est précisément la légitimité de cette

frontière qui est en jeu ici. Mais en dehors de cela, il ne semble y avoir aucune différence.

L'exemple que nous venons de donner est loin d'être aussi fantaisiste qu'il y paraît. Ce n'est pas la seule existence de ressources informatiques de pointe qui pose problème, mais la tendance générale des humains à s'appuyer sur les supports environnementaux pour raisonner. Pensez au recours à un stylo et une feuille de papier pour les longues multiplications (McClelland et al 1986, Clark 1989), au réagencement physique des jetons sur lesquels sont inscrites les lettres du Scrabble pour retrouver un mot (Kirsh 1995) et aux ustensiles que sont le langage, les livres, les diagrammes et la culture. Dans ces exemples, le cerveau de l'individu effectue certaines opérations, alors que d'autres sont déléguées aux manipulations de médias externes. Si notre cerveau était différent, la distribution de ces tâches serait sûrement différente elle aussi.

D'ailleurs, les rotations mentales décrites dans les scénarios 1 et 2 sont bien réelles. Ces exemples reflètent les options qui se présentent au joueur de Tetris. Dans ce jeu vidéo, des formes géométriques tombent à l'écran et doivent être dirigées rapidement dans les espaces appropriés au sein d'une structure émergente. Un bouton permet d'effectuer des rotations. David Kirsh et Paul Maglio (1994) ont établi que la

rotation physique d'un objet à 90° prenait environ 100 millisecondes auxquelles il faut ajouter 200 millisecondes pour choisir le bouton. Une rotation mentale demande environ 100 millisecondes pour obtenir le même résultat. Kirsh et Maglio avancent aussi des indices convaincants pour démontrer que la rotation physique est employée non seulement pour positionner une forme dans un espace, mais aussi pour *établir* la compatibilité entre la forme et l'espace. Cette utilisation constitue ce que Kirsh et Maglio appellent une « action épistémique ». Les actions épistémiques altèrent le monde pour faciliter et augmenter les processus cognitifs tels que la reconnaissance et la recherche. À l'inverse, les actions simplement *pragmatiques* altèrent le monde parce qu'une modification physique est désirable en tant que telle (par exemple utiliser du ciment pour boucher un trou dans un barrage).

Nous pensons que l'action épistémique exige une extension du *crédit épistémique*. Si, quand nous sommes confrontés à une tâche, une partie du monde fonctionne comme un processus qui, *s'il avait lieu dans notre tête*, serait reconnu sans la moindre hésitation comme une partie du processus cognitif, alors cette partie du monde *fait partie* du processus cognitif.

Le processus cognitif ne se déroule donc pas (uniquement) dans notre tête !

3. L'externalisme actif

Dans tous ces exemples, l'organisme humain est relié à une entité extérieure dans une interaction à double sens, créant un *système couplé* qui peut être vu comme un système cognitif à part entière. Tous les composants du système jouent un rôle causal actif ; ensemble, ils guident le comportement comme le fait habituellement la cognition. Si on retire les composants externes, la compétence comportementale du système s'écroule, tout comme elle le ferait si on retirait une partie du cerveau. Notre thèse est que ce type de processus couplé compte autant que le processus cognitif, qu'il se déroule entièrement dans la tête ou non.

Cet externalisme diffère grandement de la version standard proposée par Putnam (1975) et Burge (1979). Quand je crois que l'eau est mouillée et que mon double croit que le double de l'eau est mouillé, les données extérieures responsables de la différence de nos croyances sont distales et historiques, elles se trouvent à l'extrémité d'une longue chaîne causale. Les données du *présent* ne sont pas pertinentes : si je suis entouré de XYZ maintenant (j'ai peut-être été téléporté sur un double de la Terre), mes croyances portent toujours sur l'eau

standard, du fait de mon histoire. Dans ces cas-là, les données extérieures pertinentes sont *passives*. Du fait de leur nature distale, elles ne jouent aucun rôle dans la conduite du processus cognitif hic et nunc. Cela se reflète dans le fait que les actions accomplies par mon double et par moi sont physiquement indifférenciables, malgré nos différences externes.

A l'inverse, dans les cas que nous avons décrits, les données extérieures pertinentes sont *actives* et jouent un rôle crucial hic et nunc. Parce qu'elles sont couplées avec l'organisme humain, elles ont un impact direct sur cet organisme et son comportement. Dans ce cas précis, les parties du monde pertinentes sont *intégrées* et ne sont pas l'extrémité d'une longue chaîne causale. En nous concentrant sur ce type de couplage, nous aboutissons à un *externalisme actif* distinct de l'externalisme passif de Putnam et Burge.

Beaucoup ont reproché à Putnam et Burge leur manque de clarté quant à savoir si l'externalité de contenu joue un rôle causal ou explicatif dans la genèse de l'action. Dans les contre-exemples où la structure interne est maintenue alors que les données externes sont modifiées, le comportement semble identique, la structure interne semble donc responsable du travail le plus important. Nous ne nous prononcerons pas sur

cette question, mais nous remarquons que l'externalisme actif n'est pas remis en cause par un problème de ce type. Les données extérieures dans un système couplé jouent un rôle indispensable- si on conserve la structure interne en éliminant les données externes, le comportement risque de changer radicalement. Les données externes sont ici tout aussi pertinentes causalement que les traits internes typiques du cerveau.¹

Adopter un externalisme actif nous offre une explication plus simple pour toutes sortes d'actions. On peut par exemple expliquer mon choix de mot au Scrabble comme le résultat d'un processus cognitif étendu impliquant la réorganisation des lettres sur mon chevalet. Bien sûr on pourrait toujours essayer d'expliquer mon action en termes de processus internes et d'une longue série de «stimuli » et d' « actions », mais une telle explication serait inutilement complexe. Si un processus isomorphe se déroulait dans notre tête, on ne ressentirait par le

¹ Une grande partie de l'attrait de l'externalisme en philosophie vient sans doute de l'attrait intuitif de l'externalisme actif. Les externalistes font souvent des analogies impliquant des éléments externes dans des systèmes couplés, et font appel à l'arbitraire des frontières entre l'esprit et son environnement. Mais ces intuitions sont peu compatibles avec les standards de l'externalisme actif. Dans la plupart des exemples de Putman/Burge, l'environnement immédiat n'est pas pertinent, seul compte l'environnement historique. Le débat s'est concentré sur la question de savoir si l'esprit doit être dans la tête, mais une question plus pertinente serait : l'esprit se trouve-t-il dans le présent ?

besoin de le caractériser aussi laborieusement². De manière très littérale, la réorganisation des lettres sur le chevalet ne fait pas partie de l'action, mais de la *pensée*.

Cette position que nous défendons se retrouve dans un corpus de recherche toujours plus important en sciences cognitives. Dans des champs aussi divers que la théorie de la cognition située (Suchman 1987), les études de robotique (Beer 1989), les approches dynamiques du développement de l'enfant (Thelen et Smith 1994), et les recherches sur les propriétés cognitives des collectivités d'agents (Hutchins 1995), la cognition est souvent pensée en continuité avec les processus de l'environnement³. Ainsi, considérer que la cognition est étendue n'est pas seulement une décision terminologique, cela fait une différence décisive dans la méthodologie de la recherche scientifique. En effet, les méthodes explicatives que l'on pensait ne pouvoir appliquer qu'à l'analyse de processus « internes » sont désormais adaptées à l'étude de l'extérieur, et on peut espérer que notre compréhension de la cognition s'en trouvera enrichie.

² Herbert Simon (1981) a émis l'idée que nous voyions la mémoire interne comme une ressource externe sur laquelle les « vrais » processus internes opèrent. « La recherche dans la mémoire, dit-il, n'est pas très différente de celle qui opère au sein d'un environnement extérieur. » Ce point de vue avait au moins le mérite de traiter les processus internes et externes avec la parité qu'ils méritent, mais nous pensons que pour beaucoup il risquerait de trop réduire la place du cerveau.

³ On retrouve des visions philosophiques similaires dans Haugeland 1995, McClamrock 1985, Varela et al 1991, Wilson 1994.

Pour certains, cet externalisme est désagréable à accepter. Cela pourrait s'expliquer par le fait que pour beaucoup le cognitif se confond avec le conscient, et il semble alors fort peu plausible que la conscience s'étende au-delà de notre tête. Mais tous les processus cognitifs, du moins dans une conception basique, ne sont pas des processus conscients. Il est couramment accepté que toutes sortes de processus se déroulant en dehors des limites de la conscience jouent un rôle crucial dans le traitement cognitif : dans la récupération des souvenirs, les processus linguistiques, l'acquisition de compétences, notamment. Ainsi, le fait que les processus externes soient externes alors que la conscience est interne ne suffit pas à nier que ces processus sont cognitifs.

Il est plus intéressant de noter qu'un argument en faveur de l'idée selon laquelle la vraie cognition se déroule dans la tête est la nécessité que les processus cognitifs soient *portables*. Nous sommes ici transportés par une vision de ce que l'on pourrait appeler l'Esprit Nu : un ensemble de ressources et d'opérations que nous pouvons toujours appliquer à une tâche cognitive, quel que soit l'environnement immédiat. De ce point de vue, le défaut des systèmes couplés serait qu'ils sont trop facilement découplés. Les vrais processus cognitifs sont ceux

qui se situent au cœur du système : tout le reste n'est qu'un ajout.

Cette objection est intéressante. Le cerveau (ou le cerveau et le corps) comprend un ensemble de ressources cognitives fondamentales et transportables qui en elles-mêmes ne manquent pas d'intérêt. Ces ressources peuvent inclure des actions corporelles dans des processus cognitifs, comme quand on compte sur ses doigts par exemple, mais elles ne comprennent pas les aspects les plus contingents de notre environnement extérieur, comme par exemple une calculatrice. Mais la contingence du couplage n'exclut pas le statut cognitif. Dans un futur lointain nous serons sans doute capables de brancher divers modules dans notre cerveau pour nous assister dans certaines tâches : par exemple un module pour ajouter de la mémoire à court terme si nécessaire. Quand un module est branché, les processus qui y ont recours sont aussi cognitifs que s'ils avaient été présents depuis toujours.⁴

Même en faisant de la portabilité le critère central, l'externalisme actif ne serait pas remis en question. Le fait de

⁴ Ou prenez ce passage d'un roman de science-fiction récent (McHugh, 1992, p.213) : « On m'emmène au département du système auquel on me branche. Je me contente de me brancher et les techniciens lui ordonnent de se régler, ce qu'il fait. Je me débranche et demande l'heure. 10 :52. L'information apparaît. Avant quand j'avais accès à une information alors que j'étais branché, j'avais l'impression de savoir ce que je pensais et ce que le système me disait, mais maintenant comment savoir ce qui est le système et ce qui est Zhang ?

compter sur les doigts a ouvert la porte, et il est facile de pousser les choses plus loin. Pensez à l'image d'Epinal de l'ingénieur ne quittant jamais sa règle à calcul. Imaginons que les gens ne quittent jamais leur calculatrice ou qu'ils se la fassent implanter. La vraie conclusion qu'apporte l'hypothèse de la portabilité, c'est qu'un couplage *fiable* est nécessaire pour que les systèmes couplés soient au cœur de la cognition. Il se trouve que les couplages les plus fiables ont lieu à l'intérieur du cerveau, mais il peut aussi y avoir des couplages fiables avec l'environnement. Si des ressources telles que ma calculatrice ou mon Filofax sont toujours là quand j'ai besoin d'elles, alors leur couplage avec moi est aussi fiable que nécessaire. Dans les faits, elles font partie de l'ensemble de ressources cognitives que je transporte pour affecter le monde au quotidien. Ces systèmes ne peuvent être contestés simplement sur la base du risque de dommages, de perte ou de dysfonctionnement ou à cause d'un découplage occasionnel : le cerveau biologique court le même danger et il perd parfois ses capacités temporairement durant des épisodes de sommeil, d'ébriété ou d'émotion. Si les capacités pertinentes sont généralement disponibles quand on en a besoin, alors on peut parler de couplage.

De plus, il est possible que le cerveau biologique ait en fait évolué d'une manière qui tienne compte de la présence

fiable d'un environnement externe manipulable. Il semble en tout cas que l'évolution ait favorisé des capacités intégrées spécialement adaptées pour parasiter l'environnement immédiat afin de réduire le poids de la mémoire et de transformer la nature même des problèmes quantitatifs. Nos systèmes visuels ont évolué pour reposer sur leur environnement de différentes façons : ils exploitent les faits contingents sur la structure de scènes naturelles (voir Ullman et Richards 1984), et ils tirent parti de raccourcis offerts par le mouvement du corps et la locomotion (voir Blake et Yuille 1992). Peut-être existe-t-il d'autres cas où l'évolution a trouvé avantageux d'exploiter la possibilité d'intégrer l'environnement au circuit cognitif. Si tel est le cas, alors le couplage externe fait partie de l'ensemble fondamental de ressources cognitives dont nous nous servons pour peser sur le monde.

Le langage pourrait être un bon exemple, étant le moyen central par lequel les processus cognitifs sont étendus au monde. Prenez par exemple un groupe qui échange des idées autour d'une table ou une philosophe qui réfléchit mieux et développe ses idées en écrivant. On peut supposer que le langage ait évolué, en partie, pour permettre de telles extensions de nos ressources cognitives dans des systèmes activement couplés.

Au cours de la vie d'un organisme, l'apprentissage individuel peut avoir modelé le cerveau de manière à se reposer sur les extensions cognitives qui nous entouraient pendant que nous apprenions. Là encore, le langage est un exemple crucial, tout comme les divers artefacts physiques et quantitatifs utilisés quotidiennement comme extensions cognitives par les enfants à l'école et par les apprentis dans de nombreuses professions. Dans de tels cas, le cerveau se développe d'une manière qui complète les structures extérieures et apprend à jouer son rôle dans un système unifié et densément couplé. Une fois que l'on a reconnu le rôle crucial de l'environnement agissant pour contraindre l'évolution et le développement de la cognition, on voit que la cognition étendue est un processus cognitif central et non un supplément.

Une analogie pourrait aider à mieux comprendre. D'après des recherches, l'extraordinaire efficacité des poissons pour la nage est due en partie à une capacité évolutive à coupler leurs comportements avec les sources extérieures d'énergie cinétique de son environnement aqueux telles que les tourbillons, les remous et les vortex (voir Triantafyllou et G. Triantafyllou 1995). Ces vortex sont à la fois naturels (là où l'eau percute des rochers) et artificiels (créés par des battements de queue donnés au bon moment). Le poisson nage en intégrant ces processus extérieurs au cœur même de

ses mouvements de locomotion. Le poisson et les vortex environnants constituent une machine unifiée d'une efficacité remarquable.

Prenez maintenant un élément fiable de l'environnement humain, comme l'océan des mots. Nous sommes plongés dans ce bain linguistique depuis la naissance. Dans de telles conditions, le cerveau humain, plastique, en viendra sûrement à traiter ces structures comme des ressources fiables à intégrer dans l'élaboration de d'habitudes cognitives internes. Là où le poisson remue la queue pour générer les vortex et les remous qu'il exploite par la suite, nous intervenons dans de nombreux média linguistiques, et nous créons des structures locales et des agitations dont la présence fiable conduit nos processus internes continus. Les mots et les symboles externes sont cruciaux parmi les vortex cognitifs qui contribuent à construire la pensée humaine.

4. De la cognition à l'esprit

Jusqu'à présent, nous avons surtout parlé de « processus cognitif » et nous avons défendu son extension dans l'environnement. Certains pourraient penser que la conclusion a été atteinte trop facilement. Peut-être qu'une partie du processus se déroule dans l'environnement, mais quid de l'*esprit* ? Tout ce que nous avons dit jusqu'à présent est

compatible avec l'idée selon laquelle les états réellement mentaux- les expériences, croyances, désirs, émotions, etc.- sont tous déterminés par des états du cerveau. Peut-être que ce qui est vraiment mental est interne après tout ?

Nous proposons d'aller plus loin. Alors que certains états mentaux, comme les expériences, peuvent être déterminés intérieurement, il existe d'autres cas où les facteurs extérieurs apportent une contribution significative. Nous pensons en particulier que les *croyances* peuvent être constituées en partie par des données de l'environnement quand ces données jouent un rôle adéquat dans la conduite du processus cognitif. Si tel est le cas, l'esprit s'étend dans le monde.

Prenons d'abord un cas normal de croyance fondée sur la mémoire. Inga apprend par une amie qu'il y a une exposition of MoMa et décide de s'y rendre. Elle réfléchit un instant et se souvient que le musée se trouve sur la 53e Rue, elle s'y rend donc et entre dans le musée. Il semble clair qu'Inga croit que le musée est sur la 53e et qu'elle le croyait avant même de consulter sa mémoire. Ce n'était pas alors une croyance *occurrente*, mais peut de nos croyances le sont. La croyance faisait partie de sa mémoire et attendait qu'elle y accède.

Maintenant prenez Otto. Otto est atteint de la maladie d'Alzheimer, et comme beaucoup de patients, il compte sur les

informations présentes dans son environnement pour structurer son existence. Otto a toujours un carnet sur lui. Quand il apprend une nouvelle information, il la note. Quand il a besoin d'une information ancienne, il consulte son carnet. Pour Otto, le carnet joue le rôle habituellement tenu par la mémoire biologique. Aujourd'hui, Otto entend parler de l'exposition au MoMa et décide d'aller la voir. Il consulte son carnet qui lui indique que le musée se trouve sur la 53e Rue, il s'y rend donc et pénètre dans le musée.

A l'évidence, Otto s'est rendu dans la 53e Rue parce qu'il voulait aller au musée et qu'il croyait que le musée s'y trouvait. Et tout comme Inga possédait sa croyance avant de consulter sa mémoire, il semble raisonnable de dire qu'Otto croyait que le musée était sur la 53e avant de consulter son carnet. Car par bien des aspects, les deux cas sont entièrement analogues : le carnet d'Otto a le même rôle pour lui que la mémoire d'Inga. L'information contenue dans le carnet fonctionne de la même manière que l'information constituant une croyance non-occurrence ordinaire : cette information se trouve simplement au-delà du corps.

L'autre solution consiste à dire qu'Otto n'a aucune croyance sur le sujet avant de consulter son carnet : au mieux, il croit que le musée est situé à l'adresse inscrite dans son carnet.

Mais si on suit Otto pendant quelques temps, nous verrons à quel point cette manière de tourner les choses est artificielle. Otto utilise constamment son carnet. Il est crucial pour ses actions dans toutes sortes de contextes, de la même manière qu'une mémoire ordinaire est cruciale dans la vie de tous les jours. La même information peut émerger encore et encore, en étant parfois légèrement modifiée, avant de retourner dans les tréfonds de cette mémoire artificielle. On passe à côté de quelque chose quand on dit que les croyances disparaissent quand Otto range son carnet comme quand on dit que les croyances d'Inga disparaissent dès qu'elle n'en est plus consciente. Dans les deux cas, la présence de l'information est fiable et celle-ci est disponible pour la conscience et pour guider l'action, ce que l'on attend d'une croyance.

Dans la mesure où les croyances et les désirs sont caractérisés par leur rôle explicatif, les cas d'Inga et d'Otto se ressemblent : les dynamiques causales de chacun des deux exemples sont symétriques. Nous nous contentons d'expliquer les actions d'Inga par son désir occurrent d'aller au musée et sa croyance de longue date qu'il se trouve sur la 53e, et nous devrions nous satisfaire d'une telle explication pour les actions d'Otto. L'alternative consisterait à expliquer l'action d'Otto par son désir occurrent d'aller au musée, sa croyance de longue date qu'il se trouve à l'adresse notée dans son carnet et le fait

accessible que le carnet indique que le musée est sur la 53e ; mais cela complique inutilement l'explication. Si nous voulons expliquer l'action d'Otto de cette manière, alors nous devons faire de même pour toutes les innombrables actions impliquant son carnet, chaque explication comprendra alors un terme supplémentaire impliquant le carnet. Nous pensons qu'expliquer les choses ainsi reviendrait à ajouter une étape de trop. Cela serait inutilement complexe, tout comme il serait inutilement complexe d'expliquer les actions d'Inga par les croyances sur sa mémoire. Le carnet est une constante pour Otto, de la même manière que la mémoire est une constante pour Inga ; y faire allusion pour expliquer chaque croyance/désir serait redondant. Pour toute explication, la simplicité est une force.

Si tout ceci est juste, nous pouvons alors construire l'exemple d'un Otto Deux, identique au premier à ceci près qu'il a écrit dans son carnet que le musée se trouvait sur la 51e Rue. Aujourd'hui, Otto Deux est la copie conforme d'Otto, seul son carnet diffère. Ainsi, Otto Deux est caractérisé par le fait qu'il croie que le musée est sur le 51e alors qu'Otto croit qu'il est sur la 53e. Dans ces deux cas, la croyance ne se trouve simplement pas dans leur tête.

Cela rejoint les conclusions de Putman et Burge, mais il existe là aussi d'importantes différences. Dans leur réflexion, les données extérieures qui constituent des différences de croyance sont distales et historiques, les doubles produisent alors des comportements physiquement indifférenciables. Dans les cas que nous décrivons, les données extérieures pertinentes jouent un rôle actif dans le hic et nunc, et elles ont un impact direct sur le comportement : Otto se rend sur la 53e Rue, Otto Deux sur la 51e. On ne peut pas parler de manque de pertinence explicative pour ce type de contenu de croyance extérieur ; il est en question précisément du fait de son rôle central. Comme dans les cas de Putman et Burge, ces cas impliquent des différences dans la référence et dans les conditions de vérité, mais ils impliquent aussi des différences dans la dynamique de la *cognition*.⁵

La morale de tout cela, c'est qu'en termes de croyance, le crâne et la peau n'ont rien d'incontournable. Ce qui fait

⁵ Si on suit la terminologie de Chalmers « Les composants du contenu » (à paraître), les doubles des cas de Putman et Burge ne diffèrent que dans leur contenu *relationnel*, mais Otto et son double diffèrent dans leur contenu *notionnel*, le type de contenu qui règne sur la cognition. Le contenu notionnel est généralement interne à un système cognitif, mais dans ce cas le système cognitif est étendu et inclut le carnet.

d'une information une croyance, c'est le rôle qu'elle joue, et il n'y aucune raison de penser que ce rôle ne peut se jouer qu'à l'intérieur du corps.

Certains réfuteront cette conclusion. Un contradicteur pourra faire remarquer que selon son acception du terme « croyance », ou peut-être même en vertu de l'usage commun du mot, Otto ne croit pas que le musée est sur la 53e. Nous n'avons pas l'intention d'entrer dans un débat sur l'usage standard de la langue, notre propos consiste à dire que la notion de croyance *devrait* être employée pour inclure le cas d'Otto. Dans tous les aspects *importants*, le cas d'Otto est similaire à un cas normal de croyance (non-occurrence). Les différences entre Otto et Inga sont frappantes mais superficielles. En utilisant la notion de croyance dans une acception plus large, elle obtient une dimension plus naturelle. La notion est alors plus profonde et unifiée, et elle constitue une explication plus utile.

Pour proposer une réfutation efficace, il faudrait démontrer que les cas d'Otto et d'Inga diffèrent de manière importante et pertinente. Mais les deux cas sont-ils si profondément différents ? N'avancer que le fait que l'information se trouve dans la tête dans un cas et pas dans l'autre reviendrait à éluder la question. Si cette différence est

pertinente dans le cas d'une différence de croyance, elle n'est pas essentiellement pertinente. Pour justifier un traitement différent, il faut trouver une différence sous-jacente plus fondamentale.

Certains ont avancé que les cas étaient différents car Inga à un accès plus *fiable* à l'information. Après tout, n'importe qui pourrait prendre le carnet d'Otto, alors que la mémoire d'Inga est plus sûre. Il est n'est pas invraisemblable que la constance soit un élément déterminant : en effet, le fait qu'Otto utilise constamment son carnet avait contribué à justifier son statut cognitif. Si Otto consultait un guide ponctuellement, nous serions bien moins enclins à accorder à ce dernier le statut de croyance établie. Mais dans l'exemple premier, l'accès d'Otto à son carnet est très fiable, pas parfait, certes, mais la mémoire d'Inga ne l'est pas non plus. Un chirurgien pourrait lui endommager le cerveau ou, plus simplement, elle pourrait trop boire. La simple possibilité d'un tel dommage ne suffit pas à nier la croyance.

On pourrait faire remarquer que *dans les faits* l'accès d'Otto à son carnet est fluctuant. Ainsi, il se douche sans son carnet et ne peut pas le lire dans le noir. Ses croyances ne peuvent être aussi fluctuantes. On pourrait contourner ce problème en redécrivant la situation, mais dans tous les cas une

déconnexion temporaire occasionnelle n'invalide pas notre propos. Après tout, quand Inga est endormie ou ivre, on ne dit pas que sa croyance disparaît. Ce qui compte, c'est que l'information est facilement accessible quand le sujet en a besoin, et cette contrainte est également satisfaite dans les deux cas. Si le carnet d'Otto ne lui était pas accessible à des moments où les informations qu'il contient sont utiles, il y aurait un problème, car les informations ne pourraient plus jouer le rôle de guide dans l'action qui est central à la définition de croyance, mais s'il est facilement accessible dans la plupart des situations pertinentes, alors la croyance n'est pas remise en cause.

Une différence pourrait être qu'Inga a un meilleur accès à l'information qu'Otto. Les processus « centraux » d'Inga et sa mémoire partagent un lien à relativement haut débit comparé à la connexion plus faible qui unit Otto et son carnet. Mais cela ne suffit pas à différencier la croyance et la non-croyance. Imaginons une amie d'Inga, Lucy, dont la mémoire biologique serait faiblement liée à ses systèmes centraux, à cause d'un développement biologique non-standard ou de mésaventures passées. Le processus est peut-être moins efficace chez Lucy, mais tant que les informations pertinentes sont accessibles, Lucy croit que le musée est sur la 53e Rue. Si la connexion était insuffisamment directe- si Lucy avait du mal à récupérer

l'information et rencontrait des résultats inégaux ou si l'aide d'un psychothérapeute était nécessaire- nous serions moins enclins à lui attribuer une croyance, mais de tels cas sont bien différents de la situation d'Otto, pour qui l'information est accessible facilement.

On pourrait dire qu'Otto accède à l'information pertinente uniquement grâce à la perception, alors qu'Inga à un accès plus direct, par l'introspection pourrait-on dire. D'une certaine manière, présenter les choses ainsi revient à éluder la question. Après tout, nous défendons un point de vue selon lequel les processus internes d'Otto et son carnet constituent un seul et unique système cognitif. Du point de vue de ce système, le flux d'information entre le carnet et le cerveau n'est pas perceptuel : il n'implique pas l'impact de quelque chose se trouvant en dehors du système. Il est plus proche d'un flux d'information à l'intérieur du cerveau. L'accès est perceptuel dans le cas d'Otto, uniquement dans la mesure où une phénoménologie perceptuelle est associée à l'obtention de l'information, alors que ce n'est pas le cas pour Inga. Mais pourquoi la nature d'une phénoménologie associée devrait-elle faire une différence quant au statut d'une croyance ? La mémoire d'Inga peut avoir une phénoménologie associée, mais c'est tout de même une croyance. La phénoménologie n'est pas visuelle bien sûr. Mais pour ce qui est de la phénoménologie

visuelle, pensez au Terminator, dans le film éponyme avec Arnold Schwarzenegger. Quand il se souvient d'une information issue de sa mémoire, elle s'affiche dans son champ visuel (il en est sans doute conscient, puisque que de nombreux plans adoptent son point de vue). Le fait que des souvenirs soient convoqués de cette manière inhabituelle ne change pas grande chose à leur statut de croyance de longue date.

Ces quelques différences entre Otto et Inga sont *superficielles*. S'arrêter dessus reviendrait à ne pas voir comment les entrées dans le carnet d'Otto jouent le même rôle que les croyances guidant la plupart des sujets dans leur vie quotidienne.

Le sentiment que la croyance d'Otto n'est pas une vraie croyance vient sans doute de l'impression que les seules véritables croyances sont occurrentes. Si on prend ce sentiment au sérieux, la croyance d'Inga sera exclue également, comme beaucoup de croyances que nous attribuons dans la vie de tous les jours. Ce serait un point de vue extrême mais aussi le moyen le plus cohérent de nier la croyance d'Otto. D'un point de vue un tout petit peu moins radical- penser qu'une croyance doit être disponible pour la conscience par exemple- le carnet d'Otto semble tout aussi acceptable que la mémoire d'Inga.

Une fois qu'on laisse passer les croyances structurales, il est difficile de résister à la conclusion qui consiste à dire que le carnet d'Otto possède toutes les caractéristiques pertinentes.

5. Au-delà des limites externes

Si cette thèse est acceptée, jusqu'où peut-on aller ? Toutes sortes de questionnement émergent. Quid des villageois amnésiques de *Cent ans de solitude* qui oublient le nom de tous les objets et accrochent des étiquettes partout ? Est-ce que les informations contenues dans mon Filofax font partie de ma mémoire ? Si le carnet d'Otto est modifié, croit-il la nouvelle information ? Crois-je le contenu d'une page que j'ai sous les yeux avant de la lire ? Mon état cognitif est-il distribué sur Internet ?

Nous ne pensons pas qu'il existe de réponse univoque à ces questions et nous n'essaierons pas d'en proposer. Mais pour contribuer à comprendre ce qui est impliqué dans les attributions de croyance étendue, nous pouvons au moins examiner les données de notre problème central qui rend la notion si applicable ici. Tout d'abord, le carnet est une constante dans la vie d'Otto, dans les cas où les informations contenues dans le carnet sont pertinentes, il n'agira que rarement sans le consulter. Ensuite, l'information dans le carnet est disponible directement sans difficulté. Troisièmement,

quand il retrouve une information dans son carnet, il y adhère automatiquement. Enfin, les informations contenues dans le carnet ont été consciemment approuvées par le passé et c'est à la suite de cette adhésion qu'elles s'y trouvent.⁶ Le statut de ce dernier élément comme critère de croyance est discutable (on peut peut-être acquérir des croyances par une perception subliminale ou à travers une modification de la mémoire ?) mais les trois premières caractéristiques jouent un rôle crucial.

Dans la mesure où les questions de plus en plus incongrues ne tiennent pas compte de ces éléments, l'applicabilité de la notion de « croyance » diminue. Ainsi, si j'agis rarement sans consulter mon Filofax, son statut au sein de mon système sera comparable à celui du carnet d'Otto. Mais si j'agis souvent sans le consulter- par exemple si je réponds parfois « je ne sais pas » à des questions pertinentes- les informations qu'il contient occupent une place moins nette dans mon système de croyance. Internet risque de ne pas être acceptable à plusieurs niveaux, à moins que je ne sois exceptionnellement informatisé, à l'aise avec la technologie et confiant dans celle-ci, les informations présentes dans certains

⁶ Les critères de constance et d'adhésion passée pourraient laisser penser que l'histoire est en partie constitutive de la croyance. On pourrait alors éliminer tout composant historique (et ainsi donner une lecture strictement structurale du critère de constance et éliminer le critère d'adhésion passée) ou on pourrait tolérer ces composants tant que la plus grande partie du poids repose sur les éléments du présent.

dossiers de mon ordinateur pourraient alors être envisagées. Dans les cas intermédiaires, la question de savoir si une croyance est présente pourrait ne pas avoir de réponse ou la réponse pourrait dépendre de différents standards à l'œuvre dans les divers contextes où la question se pose. Mais cette indétermination ne signifie pas que la réponse n'est pas claire dans les exemples centraux.

Qu'en est-il de la cognition socialement étendue ? Mes états mentaux pourraient-ils être constitués en partie des états mentaux des autres penseurs ? En principe, nous ne voyons pas de raison de contredire cette affirmation. Au sein d'un couple exceptionnellement interdépendant, il est tout à fait possible que les croyances d'un partenaire jouent le même type de rôle pour l'autre que le carnet d'Otto⁷. Ce qui est central, c'est un haut degré de confiance et d'accessibilité. Dans d'autres relations sociales, ces critères ne seront peut-être pas aussi clairement remplis, mais ils pourront l'être dans des domaines spécifiques. Par exemple, le serveur de mon

⁷ Dans le *New York Times* du 30 mars 1995, p. B7, un article sur un ancien coach de basket universitaire, John Wooden : « Wooden et sa femme ont assisté à 36 Final Four de suite, et elle lui a toujours servi de banque de données. Nell Wooden oubliait rarement un prénom- son mari en retenait peu- et dans les salons, elle reconnaissait les gens à sa place. »

restaurant favori pourrait jouer le rôle de dépositaire de mes plats préférés (cela pourrait même être considéré comme un cas de désir étendu). Dans d'autres cas, les croyances de certains pourront être incarnées dans son/sa secrétaire, son/sa comptable, son/sa collaborateur/rice⁸.

Dans tous ces cas de figure, le plus grande partie du poids du couplage entre les agents repose sur le langage. Sans le langage, nous serions sans doute plus proches des esprits internes cartésiens discrets, dans lesquels la cognition repose surtout sur des ressources internes. Mais l'apparition du langage nous a permis de répartir ce poids sur le monde entier. C'est un outil dont le rôle est d'étendre la cognition d'une

⁸ Est-ce qu'un tel raisonnement justifie les croyances de Burge dans une arthrite étendue ? Après tout, je fais confiance à mon médecin pour prendre les mesures nécessaires pour me soigner. Certes, mais il y a des différences notables. Par exemple, des croyances étendues seraient fondées sur une relation active avec le docteur plutôt que sur une relation historique avec une communauté langagière. Et dans ce cas précis, ma relation avec le docteur donnerait plutôt quelque chose de l'ordre de la croyance vraie que j'ai une autre maladie dans ma cuisse plutôt que la fausse croyance que j'y aie de l'arthrite. D'un autre côté, si j'utilisais les experts médicaux uniquement comme consultants terminologiques, les résultats de Burge seraient comparables.

manière dont les rouages internes sont incapables. En effet, il est possible que l'explosion intellectuelle des phases récentes de l'évolution soit autant due à l'extension de la cognition permise linguistiquement qu'à un développement indépendant de nos ressources cognitives internes.

Enfin qu'en est-il du moi ? Est-ce qu'un esprit étendu implique un moi étendu ? Il semble que oui. Beaucoup d'entre nous admettent que le moi dépasse les limites de la conscience : mes croyances structurales, par exemple, constituent une partie profonde de moi. Si tel est le cas, alors ces frontières se trouvent peut-être au-delà de la peau. Les informations dans le carnet d'Otto sont une partie centrale de son identité en tant qu'agent cognitif. La conclusion est qu'Otto lui-même devrait être vu comme un système étendu, le couplage d'un organisme biologique et de ressources externes. Pour rejeter cette conclusion, il faudrait réduire la personnalité à un simple paquet d'états occurrents, ce qui menacerait l'idée de continuité psychologique. Il vaut mieux adopter une vision plus large, et considérer que les agents eux-mêmes sont distribués dans le monde.

Comme pour toute reconception de nous-mêmes, cette vision a des conséquences importantes. Elles sont évidentes pour ce qui concerne les visions philosophiques de l'esprit et la

méthodologie de la recherche en sciences cognitives, mais il y aura aussi des effets dans les domaines de la morale et du social. Ainsi, dans certains cas, interférer sur l'environnement d'une personne aura la même implication morale qu'une interférence avec cette personne. Et si on prend ceci sérieusement, certaines activités sociales devraient être perçues moins comme des communications ou des actions et plus comme de la pensée. Dans tous les cas, une fois débarrassés de l'hégémonie de la peau et du crâne, nous serons capables de nous percevoir comme de véritables créatures de ce monde.

Références

Beer, R. 1989. *Intelligence as Adaptive Behavior*. New York: Academic Press.

Blake, A. & Yuille, A. (eds) 1992. *Active Vision*. Cambridge, MA: MIT Press.

Burge, T. 1979. Individualism and the mental. *Midwest Studies in Philosophy* 4:73-122.

Clark, A. 1989. *Microcognition*. MIT Press.

Haugeland, J. 1995. Mind embodied and embedded. In (Y. Houn and J. Ho, eds.), *Mind and Cognition*. Taipei: Academia Sinica.

Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press,

Kirsh, D. 1995. The intelligent use of space. *Artificial Intelligence* 73:31-68.

Kirsh, D. & Maglio, P. 1994. On distinguishing epistemic from pragmatic action. *Cognitive Science* 18:513-49.

McClamrock, R. 1995. *Existential Cognition*. Chicago: University of Chicago Press.

McClelland, J.L, D.E. Rumelhart, & G.E. Hinton 1986. The appeal of parallel distributed processing". In (McClelland & Rumelhart, eds) *Parallel Distributed Processing, Volume 2*. Cambridge, MA: MIT Press.

McHugh, M. 1992. *China Mountain Zhang*. New York: Tom Doherty Associates.

Putnam, H. 1975. The meaning of 'meaning'. In (K. Gunderson, ed) *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press.

Simon, H. 1981. *The Sciences of the Artificial*. MIT Press.

Suchman, L. 1987. *Plans and Situated Actions*. Cambridge, UK: Cambridge University Press.

Thelen, E. & Smith, L. 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*.

Triantafyllou, M. & Triantafyllou, G. 1995. An Efficient Swimming Machine. *Scientific American* 272(3):64-70.

Ullman, S. & Richards, W. 1984. *Image Understanding*.
Norwood, NJ: Ablex.

Varela, F., Thompson, E. & Rosch, E. 1991. *The Embodied Mind*. Cambridge, MA: MIT Press.

Wilson, R. 1994. Wide computationalism. *Mind*
103:351-72.