

**A paraître dans : L. Faucher & P. Poirier (eds.) :
Philosophie & Neurosciences, Bellarmin ?**

Neurosciences et compréhension d'autrui

Joëlle Proust et Elisabeth Pacherie

Institut Jean-Nicod

CNRS-ENS-EHESS, Paris

1. Introduction

En quoi les recherches actuellement menées en neurosciences peuvent-elles être pertinentes pour la philosophie de l'esprit? Inversement, en quoi les analyses des philosophes peuvent-elles être utiles aux neurosciences? Notre propos sera d'esquisser une voie médiane entre deux attitudes extrêmes. La première, dont l'avocate la plus connue est la philosophe des sciences Patricia Smith Churchland, veut que le développement des neurosciences voue la philosophie de l'esprit et les outils conceptuels qu'elle utilise couramment aux oubliettes de l'histoire. Dans son livre manifeste, *Neurophilosophie*, paru en 1986, Patricia Churchland estimait que l'étude neuroscientifique du cerveau devait montrer le caractère radicalement erroné des catégories et schèmes conceptuels que les philosophes fonctionnalistes empruntent à la psychologie naïve. Selon elle, les neurosciences doivent conduire à une révolution conceptuelle: l'ontologie du mental sur laquelle s'appuie la philosophie de l'esprit — des catégories telles que les intentions, croyances, désirs, perceptions, souvenirs, etc. — est appelée à disparaître au profit de catégories nouvelles élaborées par les neurosciences. A l'opposé de ce matérialisme éliminativiste, on trouve une attitude, prônée notamment par Jennifer Hornsby et John McDowell, qui non seulement défend la pertinence des catégories psychologiques traditionnelles mais affirme une complète autonomie des modes d'explication qui font intervenir ces catégories par rapport aux catégories et modes d'explications qui ont cours dans les neurosciences. Dans cette perspective, les questions qui préoccupent les philosophes de l'esprit et celles qui préoccupent les neuroscientifiques appartiennent à des ordres radicalement différents. La réalité et la valeur explicative des catégories dont le philosophe fait usage ne dépendent pas de leur validation scientifique ou neuroscientifique. Les neurosciences n'ont donc pas de pertinence véritable pour la philosophie de l'esprit. Ce séparatisme s'appuie le plus souvent, implicitement ou explicitement, sur une distinction entre niveaux d'explication personnels et subpersonnels, la philosophie étant censée s'occuper d'explications de niveau personnel, tandis que les neurosciences auraient pour objet d'étude les mécanismes et processus de niveau subpersonnel. Toutefois, nombreux sont aussi les philosophes qui utilisent la distinction entre personnel et subpersonnel comme critère de démarcation des domaines d'intérêt de la philosophie et des neurosciences sans pour cela prôner une forme de séparatisme. Ils considèrent au contraire que le domaine des explications de niveau personnel n'est pas autonome mais que les phénomènes qui se manifestent au niveau personnel ne peuvent être pleinement compris que si certains faits subpersonnels sont pris en compte.

Dans un premier temps (§ 2), nous nous pencherons de plus près sur la distinction entre personnel et subpersonnel et sur l'idée selon laquelle elle permet de délimiter les domaines d'investigation de la philosophie et des neurosciences. Nous examinerons et critiquerons les arguments séparatistes qui défendent une autonomie du niveau d'explication personnel. Nous critiquerons aussi le présupposé, qui n'est pas propre aux séparatistes, selon lequel la division du travail entre philosophie et neurosciences correspond à la distinction entre niveaux personnel et subpersonnel. Dans un deuxième temps (§ 3) et afin de donner un tour plus concret à notre propos, nous prendrons pour objet d'étude la question des relations entre les théories de la mentalisation comme simulation qu'ont élaborées philosophes et psychologues au cours des vingt dernières années et les données récentes des neurosciences qui suggèrent l'existence de processus simulateurs dans le cerveau. Nous présenterons brièvement les principales thèses des théoriciens de la simulation et les principales données neurophysiologiques existantes et indiquerons les pièges que nous paraît comporter une interprétation trop directe de ces données neurobiologiques comme preuve de la validité de l'approche simulationniste de la théorie de l'esprit. Nous développerons ensuite (§ 4) une proposition sur la manière dont peut s'effectuer le passage des mécanismes simulateurs subpersonnels de bas niveau au niveau personnel de la simulation consciente. Nous ferons pour cela appel à la théorie dynamique du contrôle et à la théorie de la redescription d'Annette Karmiloff-Smith. Nous serons également amenés à distinguer deux notions de simulation et indiquerons en quoi cette distinction peut éclairer d'un jour nouveau certains des débats qui opposent avocats de la théorie de la simulation et partisans de la théorie de la théorie. Enfin, nous reviendrons en conclusion sur les leçons que l'on peut tirer de cet examen de la théorie de la simulation pour une conception des relations entre philosophie et neurosciences.

2. Niveaux personnel et subpersonnel

2.1. Enjeux philosophiques

Dan Dennett (1969) a le premier explicité la distinction entre ces niveaux d'une manière qui en révèle l'enjeu philosophique¹. La sphère du personnel recouvre ce qui est évoqué dans l'explication mentale (entendue comme l'explication fournie par le sujet conscient de son expérience ou de sa pensée) et la sphère du "subpersonnel" renvoie à l'explication neurophysiologique du phénomène:

Nous pouvons demander qu'on nous explique comment une personne retire sa main du poêle brûlant, mais nous ne pouvons pas demander d'autres explications en termes de "processus mentaux". Si nous cherchons des modes alternatifs d'explication, nous devons abandonner le niveau explicatif des personnes et de leurs sensations et activités et nous tourner vers le niveau subpersonnel des cerveaux et des événements qui se déroulent dans le système nerveux. (93)

La personne n'a accès qu'au "fait brut" de son expérience consciente", et non aux mécanismes et aux propriétés qui en sont la cause (1969, 93). Réciproquement, l'analyse subpersonnelle d'un épisode mental "change de sujet" : par exemple, l'examen de la douleur sous l'angle de ses mécanismes et de son implémentation neuronale perd de vue l'expérience subjective de la douleur. Même si les cerveaux font des discriminations associées à la douleur ressentie par le sujet, ce ne sont pas eux qui ressentent la douleur.

¹ Ludwig Wittgenstein et Gilbert Ryle ont implicitement utilisé cette distinction dans leurs attaques contre l'erreur de catégorie de ceux qui cherchent à donner des explications mécaniques de l'esprit. Cf. par exemple, Ryle (1949), pp. 18-23.

La distinction du personnel et du subpersonnel apparaît ici comme une distinction tranchée entre deux catégories d'explication: explication par les raisons du comportement des personnes et de leurs états intentionnels par opposition à une explication causale des comportements du corps et du système nerveux. C'est sur cette formulation initiale de la distinction que s'appuient le plus souvent les partisans du séparatisme. Le niveau d'explication personnel est considéré comme *autonome* et relevant de la compétence des philosophes. Les faits empiriques touchant à l'organisation du système nerveux qu'étudient les neurosciences ne sont pas censés pouvoir affecter l'application des explications de niveau personnel. Voyons comment s'organise la stratégie argumentative des séparatistes - ces philosophes qui tendent à conclure, contre l'intention de Dennett, que le point de vue personnel est le seul qui intéresse le philosophe, les neurosciences étant étrangères à ses préoccupations.

2.2. Le principe de l'autonomie de la philosophie

L'un des arguments centraux des séparatistes est tiré du contraste entre des "types d'intelligibilité" fournies par les deux modes explicatifs que nous venons de relever. Dans les termes utilisés par John McDowell, la première forme d'intelligibilité consiste à s'intéresser à ce qui doit être", l'autre "à ce qui tend à se passer".² Les explications de niveau personnel ont ceci de distinctif qu'elles (i) ont pour objectif d'expliquer les comportements de la personne en tant que telle (ii) par référence à des états intentionnels (croyances, motivations, intentions, passions) attribuables à la personne en tant que telle et qui (iii) rationalisent ce comportement. Autrement dit, elles visent à rendre ces comportements intelligibles relativement à des normes de rationalité en les situant dans un réseau d'activités rationnelles. En revanche, au niveau subpersonnel les explications font référence au cerveau et à des événements et processus cérébraux et sont de type causal.

Ces deux modes d'intelligibilité ne peuvent être mis sur le même plan. L'idéal de rationalité est en effet "constitutif", ce qui a pour effet de rendre irréductibles l'approche normative de la philosophie et l'approche descriptive de la science. Donald Davidson a souligné, dans *Mental Events*, le caractère holistique et rationnellement contraint de toute forme d'attribution mentale (à soi-même et à autrui):

Quand nous utilisons les concepts de croyance, de désir et autres concepts mentaux, nous devons nous tenir prêts, au fur et à mesure que les données empiriques s'accumulent, à ajuster notre théorie à la lumière de considérations de cohérence globale : l'idéal constitutif de la rationalité contrôle en partie chaque phase de l'évolution de ce que nous devons considérer comme une théorie en évolution". (Davidson, 1980, p. 223; trad. Engel, p. 299).

La compréhension de soi et d'autrui est articulée non par les lois causales d'une science mais par la rationalité elle-même, laquelle régit toute possibilité d'interprétation (c'est ce qui la rend "constitutive" des contenus mentaux). Comme le souligne McDowell, la forme de normativité qui est impliquée par le principe ne dépend pas de conditions particulières et des jugements hypothétiques qui seraient formées à partir d'elles (spécialité, entre autres, des sciences cognitives): elle est parfaitement générale. En outre, elle a un lien immédiat avec l'intelligibilité en général, en ce sens que si l'interprété viole la norme de rationalité, le processus d'interprétation tourne court.³

² McDowell,(1985)

³ Voir McDowell, 1998, 330-1.

Adopter le principe constitutif de rationalité ne contraint pas automatiquement à reconnaître la subjectivité essentielle du mental, comme le montre le cas de Davidson, pour qui le mental est défini par l'intentionnalité.⁴ Un certain nombre de séparatistes considèrent en revanche, dans la foulée de McDowell, que le principe constitutif conduit naturellement à l'idée que le mental est "à la fois réel et essentiellement subjectif". Pour McDowell, la subjectivité du mental⁵ recouvre non seulement l'expérience phénoménale et son contenu qualitatif; mais aussi les attitudes propositionnelles, en tant qu'elles expriment un point de vue subjectif sur le monde, et une quête individuelle d'intelligibilité des choses et des êtres. Les deux thèses (constitutivité de la rationalité, et subjectivité du mental) deviennent solidaires si les contenus mentaux eux-mêmes sont annexés à une herméneutique (à "l'espace des raisons"). Le principe constitutif de la rationalité permet alors de tirer une conséquence ontologique : quoique les concepts "sui generis" de l'intelligibilité normative subsument des éléments de la nature, leur place dans la nature n'est pas pertinente pour cette intelligibilité.⁶ Par conséquent, science et philosophie ne peuvent pas expliquer les mêmes choses - rien qui ne relève pas directement de l'attitude d'"ouverture à l'autre"- ne doit être pertinent: on ne doit pas "faire rentrer de force le mental dans un moule objectif"⁷. Comprendre le mental — ou même la nature — du point de vue du sujet (en termes personnels et de sens commun) devient le propre de la philosophie.

L'existence de deux champs de compétence pose le problème de l'efficacité causale : est-elle propre à chacun de ces champs ? Ou n'y a-t-il qu'une forme de causalité, la causalité naturelle scientifiquement comprise ? Comme le remarque Bermúdez (2000), l'un des arguments les plus répandus en faveur de l'autonomie de la philosophie consiste à voir dans les explications subpersonnelles des conditions neurophysiologiques qui réalisent les états mentaux tels qu'ils sont caractérisés au niveau personnel. C'est par exemple la théorie que développe Davidson avec son monisme anomal. Nul besoin, dans ce cas, de s'intéresser à ce niveau de pure "implémentation".

Certains séparatistes "extrémistes", comme Jennifer Hornsby, vont toutefois plus loin, en considérant que les raisons ne valent pas comme des causes au même sens où un état neurophysiologique en cause un autre. Une troisième prémisse est alors requise pour assurer la clôture du principe constitutif; cette prémisse pose l'existence d'un niveau de causalité *sui generis* pour le mental. La question typique que l'on pose à propos de l'action d'une personne est : pourquoi a-t-elle fait A? tandis que la question que l'on se pose en science est "pourquoi l'événement E s'est-il produit ?" Il n'existe pour Hornsby aucun terrain commun entre les deux explications causales, parce que, de son point de vue, "vouloir" ou "croire" ne peuvent pas être pensés comme des événements. Il existe ainsi un "écart" (*gap*) entre l'état mental de l'agent, d'un côté, et les événements neurophysiologiques qui se produisent dans le cerveau de l'agent, de l'autre; cet écart n'importe pas pour l'agent (qui est conscient de son existence), puisqu'il se comble avec l'effectuation réussie de l'action.⁸

⁴ Davidson écrit : "La marque distinctive du mental n'est pas son caractère privé, subjectif, ou immatériel, mais le fait qu'il manifeste ce que Brentano appelait de l'intentionnalité." (Davidson, 1980, p. 211; trad. Engel, p. 282).

⁵ Thomas Nagel (1986) est évidemment l'auteur qui a exploré le plus attentivement les différences de méthodes et les enjeux des conceptions objectives et subjectives du mental.

⁶ McDowell, 1994, 74-75.

⁷ McDowell, 1998, 336

⁸ Hornsby, 1993, 167.

2.3. Contre l'autonomie

L'argumentation séparatiste a suscité diverses réponses dans le camp adverse. On peut tout d'abord contester que l'interprétation séparatiste constitue la seule lecture possible de la distinction entre personnel et subpersonnel. Dennett (1969) soulignait que l'introduction d'une distinction entre niveaux d'explication amenait avec elle l'obligation de relier ces niveaux et que cette obligation incombait essentiellement aux philosophes. On peut admettre qu'il existe une distinction importante entre explications de niveau personnel et explications de niveaux subpersonnel sans en faire une distinction rigide entre niveaux d'explication autonomes délimitant les sphères de compétence respectives de la philosophie et des neurosciences.

En outre, la conception de l'autonomie explicative du niveau personnel que défendent les séparatistes repose sur une vision très idéalisée de la rationalité humaine, prenant pour modèle un agent parfaitement rationnel et dont les comportements et états de niveau personnel sont toujours susceptibles d'explications rationalisantes. Le séparatisme présuppose également que les normes de rationalité auxquelles sont soumises les explications de niveau personnel ont un caractère a priori et une validité universelle. Or, ni ce présupposé sur la nature des normes rationnelles, ni l'idée que le modèle d'un agent idéal parfaitement rationnel constitue un modèle valide pour la compréhension des comportements et états mentaux de niveau personnel des êtres humains ne peuvent être tenus pour des vérités intangibles.

Une question cruciale concerne ici le recours à l'intuition en matière de normes rationnelles. Comme le rappelle Stich (2003), la méthode philosophique traditionnelle de validation d'une norme rationnelle consiste à examiner des affirmations normatives en les confrontant aux jugements spontanés de tout un chacun sur des cas réels ou hypothétiques. Elles présuppose l'idée d'une sorte d'intuition rationnelle à l'œuvre dans ces jugements spontanés, intuition qui manifesterait une forme de connaissance *a priori* des normes de rationalité, soit que, comme le voulait Platon, elle nous donne accès à des Formes idéales, soit que, dans une version contemporaine laïcisée telle que la théorie de l'équilibre réfléchi, l'intuition constitue le seul mode de justification possible des règles inférentielles avec lesquelles nous opérons. Cette méthode présuppose également le caractère universellement partagé de ces intuitions. Or, selon Stich, les travaux récents de psychologues cognitivistes spécialisés dans la diversité culturelle viennent remettre en question cette vénérable méthode philosophique en mettant en évidence la variabilité culturelle de ces intuitions.

Les travaux récents de Nisbett et de ses collègues ont montré d'importantes différences entre les modes de pensée des individus de culture chinoise et ceux des occidentaux. Ces différences se manifestent de manière systématique dans une longue liste de processus cognitifs comprenant notamment l'attention, la mémoire et la perception. Asiatiques et occidentaux diffèrent aussi dans leur manière de décrire, prédire et expliquer les événements, dans leur façon de catégoriser des objets et dans leur manière de réviser leurs croyances face à de nouveaux arguments et de nouvelles données (Nisbett 2002).

S'inspirant de ces travaux, Stich a cherché à savoir s'il pouvait aussi y avoir des différences entre les intuitions sur ce qu'*est* la connaissance. Avec ses collègues (Weinberg et al., 2001), il a entrepris de tester un ensemble « de supports d'intuition » [*intuition probes*] d'ordre philosophique auprès de divers groupes. Ils ont trouvé un grand nombre de supports d'intuition – tous sur le modèle des cas hypothétiques qui ont été largement discutés par les épistémologues (tels que les exemples de Gettier) – sur lesquels différents groupes ont des intuitions significativement différentes. Dans certains cas, les différences existent entre personnes de milieux culturels différents et dans d'autres elles existent entre personnes ayant un statut socio-économique différent. Ainsi les Américains de statut socio-économique élevé

ont-ils des intuitions épistémiques différentes des Américains de niveau socio-économique faible, différences qui de surcroît sont souvent considérables. Si nos intuitions épistémiques sont affectées par des facteurs culturels et socio-économiques, l'idée de normes rationnelles a priori universelles paraît menacée et celle d'un agent rationnel idéal a des parfums d'ethnocentrisme. Comme le fait remarquer Stich, les soi-disant normes a priori de rationalité s'appuient largement sur les intuitions épistémiques d'occidentaux de haut niveau socio-économique!

Confronté à ces travaux, le partisan du séparatisme pourra répondre qu'ils remettent certes en cause l'autonomie du niveau d'explication personnel vis-à-vis d'explications faisant intervenir des facteurs culturels et socio-économiques, mais non son autonomie relativement à des explications de niveau subpersonnel. On doit noter toutefois que les sujets n'ont en règle générale nulle conscience des influences de leur culture ou de leur niveau socio-économique sur leurs intuitions. Cette influence s'exerce donc probablement via des mécanismes subpersonnels.

On peut aussi opposer des arguments plus directs à l'idée d'une autonomie des explications de niveau personnel par rapport aux explications de niveau personnel. Le séparatisme a pour thèse centrale que ce qui fait du niveau personnel un niveau d'explication autonome est le rôle qu'y joue la rationalité. De surcroît, cette rationalité est conçue sur le modèle de la rationalité logique; autrement dit, les principes inférentiels qui gouvernent les procédures de prédiction et d'explication au niveau personnel sont censés être les principes de la logique déductive et de la théorie de la probabilité. Certes, les séparatistes ne prétendent pas que dans les faits les êtres humains soient idéalement rationnels. Ils estiment toutefois que ces normes inférentielles représentent les idéaux de rationalité auxquelles nous aspirons : même lorsqu'un agent échoue manifestement à les satisfaire, on peut néanmoins comprendre son comportement en supposant qu'il s'efforce de les respecter. Or, cette thèse ne rend pas compte du fait que non seulement les êtres humains sont loin d'être idéalement rationnels, mais que les écarts que leur comportement manifeste vis-à-vis des principes de rationalité logique ne sont pas aléatoires : ils présentent un caractère systématique, mis en évidence dans toute une série d'études expérimentales.⁹ Ce que ces travaux expérimentaux montrent c'est qu'il est erroné de concevoir les stratégies de raisonnement que nous mettons en œuvre dans la vie de tous les jours comme des applications imparfaites des techniques inférentielles prescrites par la théorie normative de la rationalité. Les principes inférentiels qui gouvernent le raisonnement humain dans la vie quotidienne ne sont tout simplement pas assimilables aux principes inférentiels de la logique déductive ou du calcul des probabilités. Du coup, l'autonomie du niveau d'explication personnel se trouve menacée. Les principes de la théorie normative de la rationalité peuvent encore avoir une valeur prescriptive et être utilisée de manière réflexive par des agents rationnels cherchant à évaluer et contrôler leurs propres délibérations, mais il ne s'ensuit pas que ces principes guident nos stratégies spontanées de raisonnement et de prise de décision. Ils ne peuvent donc être utilisés à des fins d'explication et de prédiction de nos comportements ordinaires. Pour comprendre ceux-ci, il faut que nous sachions quels sont les principes inférentiels dont nous faisons tacitement usage. Or comprendre ces principes revient à comprendre des faits subpersonnels. En d'autres termes, il n'est pas toujours possible de donner des faits de niveau personnel des explications se situant entièrement au niveau personnel.

L'autonomie du niveau d'explication personnel se trouve donc remise en cause à la fois par le haut et par le bas. Par le haut dans la mesure où les normes de rationalité auxquelles nous adhérons explicitement sont influencées par des facteurs culturels et socio-économiques. Par

⁹ Sur ce point, cf. Bermudez, (2000).

le bas, dans la mesure où les principes inférentiels dont nous faisons tacitement usage dans les raisonnements de tous les jours ne correspondent pas aux principes dégagés par la théorie normative de la rationalité.

Les séparatistes pourraient être tentés d'adopter une position de repli ; plutôt que d'affirmer que la marque distinctive des explications de niveau personnel est le rôle qu'y joue la rationalité entendue au sens de la théorie normative, ils pourraient dire qu'il s'agit du seul niveau où ont cours des explications faisant appel à des états intentionnels et où interviennent des généralisations qui font référence au contenu de ces états. La caractéristique propre aux explications de niveau personnel serait alors de constituer un niveau d'explication sémantique, alors que les explications de niveau subpersonnel seraient au mieux syntaxiques et computationnelles. Mais cette conception sémantique de l'autonomie du niveau d'explication personnelle se trouve remise en cause dans nombre de modèles développés en psychologie cognitive où des états subpersonnels sont traités, au moins implicitement, comme dotés d'un contenu intentionnel. C'est le cas par exemple des modèles de la perception visuelle qui expliquent comment certaines propriétés intentionnelles de notre expérience visuelle peuvent être comprises comme le résultat de certaines opérations computationnelles sur le contenu d'états subpersonnels du système visuel. C'est le cas encore de modèles expliquant comment nous comprenons une phrase comme dotée d'une certaine signification. Peacocke (1994), par exemple, s'est efforcé d'expliquer et de justifier ce type de pratique, en montrant à la fois qu'une telle stratégie d'explication computationnelle n'échappe à l'incohérence que si l'on renonce à une conception purement syntaxique de la computation au profit d'une conception sémantique et qu'une théorie sémantique de la computation est tout à fait légitime. La stratégie argumentative de Peacocke est relativement complexe. Peacocke s'appuie sur une conception externaliste du contenu intentionnel. Il soutient qu'une explication de niveau subpersonnel des propriétés intentionnelles et donc relationnelles de niveau personnel ne peut être satisfaisante que si elle fait référence à des états computationnels subpersonnels eux-mêmes relationnels. Il est à son avis légitime d'attribuer des contenus aux états subpersonnels dans la mesure où l'on peut énoncer les principes gouvernant l'attribution correcte de contenus à des états computationnels subpersonnels. Pour autant que Peacocke et d'autres philosophes qui ont argumenté dans le même sens aient effectivement montré qu'une conception sémantique des computations opérant sur les états subpersonnels est légitime, la notion d'explication sémantique ne peut plus être tenue pour équivalente à la notion d'explication de niveau personnel.

2.4. Problèmes de frontière

La caractérisation que Dennett avait initialement proposée de la distinction entre niveaux personnel et subpersonnel et qui supposait une solidarité entre états et comportements de niveau personnel, états et comportements intentionnels et explication de niveau personnel, a été mise à mal par ce type de réflexions. S'il n'est pas toujours possible de rendre pleinement compte de faits de niveau personnel par des explications de niveau personnel, la distinction du personnel et du subpersonnel ne peut plus être conçue comme une distinction catégoriale entre deux styles d'explication mutuellement exclusifs. S'il est légitime d'attribuer un contenu représentationnel à certains états subpersonnels, les explications s'appuyant sur des généralisations qui font référence au contenu d'états représentationnels ne peuvent plus automatiquement être tenues pour des explications de niveau personnel. En conséquence, la distinction du personnel et du subpersonnel telle qu'elle est souvent utilisée aujourd'hui met moins l'accent sur la différence des styles d'explication que sur la différence des *explananda*. D'une distinction entre *explications* de niveau personnel ou subpersonnel, on est passé à une distinction entre *états et processus* personnels et subpersonnels. Comme d'autre part, il est

aussi devenu courant d'attribuer un contenu représentationnel à des états subpersonnels et de considérer que l'opération d'au moins certains processus subpersonnels est fonction du contenu des états sur lesquels ils opèrent, la distinction entre faits personnels et faits subpersonnels ne peut plus s'appuyer sur le critère du contenu intentionnel. Il semble que le critère qui prévaut aujourd'hui soit un critère d'accessibilité à la conscience. Sont considérés comme de niveau personnel les états et processus conscients ou accessibles à la conscience et comme subpersonnels les états et processus qui ne le sont pas.

Toutefois, un tel critère ne permet pas toujours d'opérer une distinction nette du personnel et du subpersonnel. La frontière, si frontière il y a, est labile et dynamique. Des procédures qui font l'objet d'une acquisition explicite et dont nous contrôlons d'abord consciemment l'exécution (le coup de revers lorsqu'on apprend à jouer au tennis, le changement de vitesse lorsqu'on apprend à conduire, etc.) peuvent avec le temps et la pratique s'automatiser et ainsi se déclencher et se dérouler en dehors du contrôle conscient. Inversement, des processus normalement inconscients peuvent dans certaines conditions devenir accessibles à la conscience. Ainsi, lors d'un mouvement volontaire, les processus de correction et d'ajustement qui sont normalement automatiques et inconscients peuvent devenir conscients lorsque la différence entre la trajectoire prévue et la trajectoire effectivement réalisée par l'effecteur excède un certain seuil. Plus généralement, comme nous le verrons plus en détail dans la section 4, au cours du développement cognitif des procédures initialement implicites sont réencodées et transformées en connaissances explicites (Karmiloff-Smith, 1992).

La démarcation entre personnel et subpersonnel entendue comme distinction entre ce qui est accessible à la conscience et ce qui ne l'est pas est donc relativement fluctuante. Permet-elle encore cependant de délimiter les sphères de compétence respectives de la philosophie et des neurosciences? Le doute est permis. Certes, le domaine d'intérêt premier des philosophes concerne l'explication des phénomènes personnels, mais les considérations dont nous avons fait état dans cette section donnent à penser que les phénomènes de niveau personnel ne constituent pas un niveau d'explication autonome et que l'on ne peut en donner d'explication satisfaisante sans prendre en compte des faits subpersonnels. En outre, comme le soulignait Dennett, comprendre la relation entre niveaux personnels et subpersonnels est une tâche à laquelle le philosophe ne peut se soustraire.

Inversement, c'est aussi avoir une vision périmée des neurosciences que de vouloir cantonner leur activité à l'explication des phénomènes de niveau subpersonnel. Certes, pendant longtemps, la neurophysiologie — la branche des neurosciences qui étudie l'organisation fonctionnelle du cerveau et ses relations avec l'activité mentale et le comportement — est restée marquée par une forte tradition réductionniste et s'est concentrée sur des problèmes jugés relativement simples tels que l'organisation des systèmes sensoriels ou la programmation des mouvements des différents effecteurs. L'étude des fonctions cognitives supérieures apparaissait comme un idéal encore inaccessible. Processus mentaux et processus cérébraux étaient bien considérés comme en principe liés, mais l'intégration des recherches sur les processus mentaux et des recherches sur le cerveau semblait en pratique irréalisable, ou en tout cas prématurée. L'émergence depuis la fin des années 1980 des neurosciences cognitives témoigne d'un changement d'attitude: l'étude des relations entre processus mentaux et cérébraux apparaît non seulement comme possible mais comme essentielle à de nouveaux progrès. L'action, l'intention, la cognition sociale ou la conscience constituent quelques-uns des domaines qui ont le plus bénéficié de ce nouvel état d'esprit. Ainsi, dans le domaine de l'action, les efforts faits pour interpréter les données anatomiques et physiologiques à l'aide de théories et de méthodes développées en psychologie cognitive, l'utilisation des données issues de la neuropsychologie et de la neuro-imagerie fonctionnelle pour tester les modèles cognitifs de la génération de l'action, la mise en relation des données

neurophysiologiques, cognitives et comportementales chez les sujets normaux, les sujets cérébro-lésés et les sujets souffrant de psychopathologies ayant trait à l'action ont donné lieu à de multiples découvertes. Elles ont mis en évidence la hiérarchie complexe des représentations et processus qui interviennent dans la préparation et le contrôle de l'action et ont permis l'élaboration d'hypothèses précises sur la structure cognitive et les bases neurales de ces processus et représentations et sur les liens entre les aspects subpersonnels de la préparation et du contrôle moteurs et des phénomènes personnels tels que la conscience du mouvement ou le sens de l'agentivité.

Il est donc clair que les neurosciences telles qu'elles se pratiquent aujourd'hui ne considèrent plus que l'explication de faits de niveau personnel soit hors de leur portée ou de leur sphère légitime d'intérêt. Du coup, les philosophes peuvent se sentir dépossédés. S'ils maintiennent, contre vents et marées, que leur domaine de compétence est constitué par les phénomènes de niveau personnel susceptibles d'explications entièrement situées au niveau personnel, ils sont condamnés à voir celui-ci se réduire à une peau de chagrin alors que s'accumulent les données empiriques montrant le rôle de facteurs subpersonnels dans les explications de phénomènes personnels. S'ils affirment simplement que leur domaine d'intérêt premier concerne les faits de niveau personnel sans présupposer de surcroît qu'à ces faits correspond un mode unique et distinctif d'explication rationnelle, ils entrent en concurrence avec les neurosciences et ne peuvent plus prétendre que la philosophie de l'esprit a un objet d'étude qui lui est propre. Plutôt que de chercher maintenant à définir dans l'abstrait ce qui est le propre de l'activité philosophique lorsqu'elle se penche sur le domaine de la cognition, nous allons tenter de suivre l'injonction de Dennett pour qui il incombe au philosophe de tenter de comprendre comment relier différents niveaux d'explication. Nous prendrons donc pour cible la notion de simulation et l'utilisation qui en est faite en philosophie et dans les neurosciences.

3. La simulation relève-t-elle d'un niveau personnel ou subpersonnel ?

3.1. Simulation et mentalisation

La "théorie de la simulation" propose d'expliquer la manière dont on comprend autrui et prédit ce qu'il fera ou ressentira par la projection imaginative du sujet dans la situation ou dans l'état mental d'autrui. Le rôle de la simulation est tantôt conçu comme suffisant à l'acquisition de concepts mentaux (Robert Gordon), tantôt comme une des voies d'accès à la mentalisation (Alvin Goldman), qui n'exclut pas la nécessité d'une théorisation mentaliste, ou d'un traitement modulaire spécialisé. La théorie de la théorie, développée entre autres par H. Wimmer et J. Perner, et A. Gopnik, part de l'hypothèse qu'un raisonnement factuel appliquant les concepts de la psychologie ordinaire est nécessaire pour comprendre autrui comme sujet intentionnel. Une autre théorie, dite "modulaire", développée par Alan Leslie (Leslie, 1987), considère que la mentalisation dépend de la maturation d'un module inné d'extraction et de traitement de l'information : c'est la capacité de découpler deux représentations de la même situation, manifestée par l'enfant de deux ans - lorsqu'il fait semblant d'être un indien etc.- qui est cruciale pour attribuer des états mentaux à un sujet (soi-même ou autrui). Le théoricien de la simulation estime de son côté que la compréhension mentaliste d'autrui dépend au moins en partie d'un processus non doxastique (non fondé sur l'attribution de croyances) : pour attribuer un état mental à autrui, le sujet doit se mettre lui-même en résonance avec les états mentaux de la cible, et tirer de cette simulation la décision, l'état émotionnel ou épistémique qu'il s'agit de comprendre ou de prédire. Dans cet usage "empathique", la simulation est présentée comme une stratégie de niveau personnel. Les philosophes décrivent

le processus comme *l'imagination* par le sujet de ce que pense ou ressent autrui dans un contexte déterminé. Certains psychologues ont testé diverses prédictions de la théorie, dans la mesure où elle suppose l'intervention de raisonnements sur des situations contrefactuelles; ils ont mis en évidence, en particulier, les différences de raisonnement selon qu'il concerne la seule situation imaginée ou la réalité (Dias et Harris, 1988, Harris et Kavanaugh, 1993) et l'intervention des procédures de « montée » (ascent routines) qui permettent, selon Gordon, de passer de l'observation d'un contexte à l'attribution épistémique correspondante (Perner, Baker & Hutton, 1994). Plus récemment, dans le cadre des théories hybrides de la mentalisation, on s'est intéressé aux conditions qui peuvent favoriser l'emploi d'une simulation plutôt que d'une conceptualisation de type théorique (Robinson et Mitchell, 1995, Perner et al. 1999, Perner et Kühberger, 2003). Ces nouvelles approches vont de pair avec l'idée que la stratégie adoptée dépend non du jugement conscient du sujet, mais des caractéristiques de la tâche de raisonnement. Comme nous le verrons plus loin, il y a ici l'amorce d'un "tournant subpersonnel" dans l'application de la théorie de la simulation à la mentalisation.

3.2. La simulation dans les neurosciences

De leur côté, les neurophysiologistes ont commencé depuis une vingtaine d'années à élucider les bases cellulaires de la compréhension des actions et émotions d'autrui. Les techniques électrophysiologiques d'enregistrement unicellulaire chez le singe éveillé ont notamment permis l'identification de deux régions cérébrales, temporale et prémotrice, sélectivement impliquées dans la perception des émotions et actions d'autrui. David Perrett et ses collaborateurs de l'Université de St Andrews ont mis en évidence au sein du sillon temporal supérieur (STS) chez le singe macaque des populations de neurones impliqués dans le codage des postures corporelles, des mouvements biologiques, des actions finalisées et enfin de la distinction entre mouvement auto-généré et mouvement produit par autrui (Perrett, 1999). L'équipe de Giacomo Rizzolatti à Parme a quant à elle découvert dans la partie rostrale du cortex prémoteur central (aire F5) du singe macaque deux classes de neurones aux propriétés remarquables (Rizzolatti et al., 1988; Gallese *et al.*, 1996 ; Fogassi et Gallese, 2002)). Une grande partie des neurones de F5, les neurones canoniques, répondent à la présentation visuelle d'objets de taille et de forme différentes en l'absence de tout mouvement détectable. La décharge de ces neurones ne dépend pas nécessairement de l'exécution d'une action mais code la représentation motrice d'une action adaptée à une interaction correcte avec l'objet. Les neurones identifient donc ces objets non pas en termes de leurs caractéristiques visuelles mais en termes moteurs. Les neurones de la deuxième classe, appelés neurones miroirs, sont activés aussi bien lorsque l'animal réalise une action donnée que lorsqu'il observe l'expérimentateur ou un autre animal exécutant la même action. Ils partagent les propriétés motrices des neurones canoniques mais s'en distinguent par leurs propriétés visuelles. Ni la vision d'un objet seul, ni celle d'un agent seul mimant une action ne suffit à les activer. La majorité de ces neurones répondent à l'observation d'un seul type d'action.

Les modes d'activation des neurones canoniques et des neurones miroirs suggèrent qu'une action donnée peut être représentée par le système nerveux indépendamment de sa transformation en une action réelle. Les deux classes de neurones partagent un même « vocabulaire » moteur mais se distinguent par leur mode d'accès à ce vocabulaire: à partir de la vision d'objets pour les neurones canoniques et à partir de l'observation d'actions exécutées par autrui pour les neurones miroirs. En outre, les neurones miroirs paraissent constituer un

système de mise en correspondance entre observation et exécution d'action, le lien entre observation et exécution étant constitué par la présence dans les deux cas d'un but¹⁰.

Étant donné la présence dans deux régions cérébrales distinctes, le sillon temporal supérieur et l'aire prémotrice ventrale de neurones dotés de propriétés visuelles complexes similaires, on peut s'interroger sur leurs possibles relations. Tant Perrett (1999) que Gallese et Goldman (1998) font l'hypothèse que l'analyse visuelle des actions est accomplie initialement dans le cortex temporal et que la sensibilité visuelle aux actions dans le cortex prémoteur dépend d'informations en provenance des cellules de codage de l'action dans le cortex temporal. Gallese et Goldman font en outre l'hypothèse que le rôle des cellules de F5 serait de conférer une signification aux descriptions visuelles des actions fournies par les neurones du STS; ils mettent ainsi l'accent sur le rôle de l'action dans l'interprétation de ce qui est perçu.

On ne peut utiliser chez les êtres humains les techniques d'enregistrements unicellulaires utilisées chez les singes macaques. Il existe toutefois des données issues des techniques d'imagerie qui suggèrent fortement l'existence chez l'humain de populations de cellules ayant des propriétés fonctionnelles semblables à celles des populations qui viennent être décrites, y compris un système de correspondance entre exécution et observation similaire au systèmes des neurones miroirs (Fadiga *et al.*, 1995; Rizzolatti *et al.*, 1996; Buccino *et al.*, 2001; Rizzolatti *et al.*, 2002). En outre, au cours de la dernière décennie, toute une série d'études de neuroimagerie ont montré l'existence d'un important recouvrement des régions cérébrales impliquées dans la génération de l'action, la simulation mentale de l'action et la perception des actions accomplies par d'autres agents (Decety *et al.*, 1994; Grafton *et al.*, 1996; Grèzes and Decety, 2001; Rizzolatti *et al.*, 1996; Stephan *et al.*, 1995). L'étude de patients neurologiques montre en outre que des déficits de la performance motrice peuvent se refléter au niveau de la simulation motrice (Dominey *et al.*, 1995; Sirigu *et al.*, 1996).

Certains ont vu dans ces découvertes la confirmation par la neurobiologie de l'approche simulationniste de la théorie de l'esprit développée par des philosophes et psychologues. Selon Perrett (1999) le traitement visuel de haut niveau effectué dans le STS permettrait la reconnaissance des émotions d'autrui à travers l'identification de leur expression faciale. Il permettrait également la compréhension de la direction de l'attention d'autrui et ainsi le suivi du regard nécessaire à l'attention partagée. Enfin, il permettrait la reconnaissance de la finalité des actions observées à travers l'identification des composantes du mouvement et l'identification de leur but.

Les populations de cellules étudiées dans le cortex prémoteur et en particulier le système de correspondance entre exécution et observation réalisé par les neurones miroirs pourraient constituer le fondement d'un ensemble de comportements jouant un rôle essentiel dans la cognition sociale. Selon plusieurs auteurs (Gallese et Goldman, 1998; Arbib et Rizzolatti, 1997; Arbib 2002) ce système de correspondance interviendrait dans la reconnaissance, l'apprentissage par imitation, l'imitation, la compréhension des actions d'autrui et pourrait être une condition nécessaire de l'évolution des capacités linguistiques. Selon Gallese et Goldman, le système des neurones miroirs pourrait même avoir une fonction beaucoup plus générale, en étant à la base non seulement de la simulation de l'action mais de la simulation des états mentaux d'autrui.

¹⁰ Pour une présentation détaillée des recherches sur les neurones miroirs et de leurs implications pour la compréhension de l'action, le développement de la cognition sociale et de la communication, voir Stamenov & Gallese (2002).

Il importe toutefois de ne pas se lancer dans des spéculations débridées et d'interpréter ces données avec prudence. Trois séries de remarques nous permettront de tenter de circonscrire les difficultés. Premièrement, on doit souligner que si ces diverses populations de cellules peuvent constituer des conditions nécessaires des différentes formes de cognition sociale qui viennent d'être énumérées, rien ne permet de conclure qu'elles en constituent des conditions suffisantes. Le singe macaque chez qui ces populations de neurones ont été étudiées n'est connu ni pour ses capacités d'imitation, ni pour ses capacités d'attention conjointe et moins encore pour ses capacités de mentalisation.

Deuxièmement, lorsque l'on affirme que ces neurones pourraient constituer les bases neurales de la compréhension d'autrui, il importe, comme y insiste Perrett, de distinguer deux interprétations possibles de l'expression "compréhension d'autrui": dans l'interprétation *mentaliste*, comprendre autrui revient à lui attribuer des états mentaux permettant de rendre compte de son comportement ; dans l'interprétation *comportementale*, l'animal apprend à associer des indices comportementaux pour prédire les réactions de ses congénères, sans faire référence à des états mentaux intermédiaires. L'un des problèmes intéressants que pose cette distinction (qui oppose la cognition sociale des primates non-humains et humains) est de savoir si elle coïncide avec la distinction entre deux capacités humaines : la compréhension pratique, tacite, d'autrui qui se manifeste dans nos réponses comportementales aux comportements que nous observons chez autrui et la compréhension théorique manifestée par la construction de représentations conscientes des états mentaux d'autrui.

Quelle que soit la réponse à ce problème, on ne peut conclure, du fait que des représentations subpersonnelles des émotions ou des actions d'autrui sont activées, à la présence de représentations personnelles conscientes. Ainsi que le souligne Arbib (2002) parlant des neurones canoniques et des neurones miroirs, lorsque l'on dit d'un individu qui exécute ou observe une action qu'il "sait" de quelle action il s'agit, il importe de bien distinguer la connaissance de l'action entendue comme le fait d'avoir une représentation neurale du mouvement et de son but, de la connaissance de l'action comme représentation consciente que peut avoir un agent de ce qu'il est en train de faire. Au niveau du système des neurones miroirs c'est seulement au premier de ces deux sens que l'on peut parler de connaissance de l'action. Certes le système a traité l'information motrice pertinente, mais ce n'est là qu'une condition nécessaire (et non suffisante) de l'accès à la connaissance correspondante.

Troisièmement, de la simulation de l'action à la simulation des états mentaux en général, le pas à franchir reste considérable. Il paraît certes légitime de décrire l'activité du système des neurones miroirs comme une forme de simulation de l'action dans la mesure où d'une part l'observateur utilise ses propres états pour mimer ou répliquer l'état dans lequel se trouve l'agent observé (la représentation de l'action motrice activée chez celui-ci) et où d'autre part les conséquences motrices de l'action observée sont inhibées chez l'observateur. Toutefois, le système des neurones miroirs ne concerne que la perception des actions (plus exactement, des actions qui impliquent la face et les mains). En outre, la simulation mentalisatrice ne peut se contenter de simuler les actions habituelles des congénères, ni rester limitée à simuler la dynamique prévisible des gestes ; elle doit pouvoir simuler des actions inhabituelles et inappropriées pour le succès de l'action dans la réalité (pour expliquer, en particulier, la réussite du simulateur dans les tâches de "croyance fausse"). La généralisation du rôle de la simulation - de l'action à l'ensemble des états mentaux - doit ainsi être justifiée.

3.3. Une transition trop rapide

Robert Gordon¹¹ propose une telle justification : les neurones-miroirs contribuent à élaborer ce qu'il appelle une "analyse par synthèse". Le cerveau, résonant à l'action observée, cherche à la rendre "non surprenante". Pour comprendre autrui, il suffit d'assigner à l'agent l'interprétation intentionnelle qui correspond aux réponses endogènes (celles qui sont suscitées par la mise en résonance). L'opération des neurones-miroir permet alors "au cerveau" de comprendre ("par défaut") les raisons d'agir et les buts d'autrui, à partir des siens propres. Gordon commente ainsi ce processus : « Le cerveau semble chercher un correspondant endogène à l'intrus exogène ». Comment alors distinguer mes actes et mes intentions d'agir de ceux d'autrui ? Supposons que je voie mon fils jouer au football. J'anticipe par les neurones miroir ce que son corps va faire, en me plaçant d'un point de vue égocentrique, et dans le contexte du raisonnement qui est le sien; j'ai une phénoménologie particulière. Tout cela contribue à former l'impression de comprendre l'action d'autrui, et non l'impression d'agir moi-même.

L'objectif de Gordon est de montrer que le système miroir réalise un niveau implicite de reconnaissance des agents intentionnels qui ne suppose aucune inférence ni raisonnement par analogie. Ce qui est remarquable dans l'explication offerte est non sa nouveauté,¹² mais l'interprétation *personnelle* qui est donnée des enregistrements de neurones individuels. Or en interprétant directement l'activité cérébrale du cerveau en termes intentionnels, on rend certes intelligible l'opération du cerveau en "traduisant" son activité dans le langage de la psychologie ordinaire. Mais à quel prix ? Et quelle intelligibilité a-t-on finalement obtenu ?

Le prix à payer est la circularité : on a *ipso facto* réinséré l'*explanandum* dans le schéma explicatif. Et de quel type est l'intelligibilité obtenue ? La psychologie ordinaire du sujet devient la psychologie du cerveau; celui-ci a des intentions, des raisons, il cherche à comprendre, à diminuer la surprise; il empathise, etc.. En outre, les explications neuronales et les explications personnelles deviennent inextricablement mêlées. Comment savoir que les actions observées sont celles d'autrui, et non les miennes ? En reconnaissant que les raisons d'agir (auxquelles je "résonne") ne sont pas les miennes. Mais comment le "Je" peut-il coopérer avec son propre cerveau pour lui communiquer les informations requises ?

L'"analyse par synthèse", ainsi comprise et appliquée, est donc un piège de l'intuition; elle consiste à aborder un problème d'analyse neurocognitive avec les moyens de la logique ordinaire ("folk logic"). L'expression revient à prêter au cerveau le type d'accès intuitif qu'a un sujet qui recompose consciemment les éléments de ce qu'il doit comprendre. Appliquée à ce que fait le cerveau, l'expression est une métaphore; elle ne peut être véritablement éclairante que si le cerveau censé faire l'analyse a la capacité d'extraire et de traiter l'information pertinente pour achever l'analyse dans ses moindres détails ; dissocier l'endogène de l'exogène, le contexte essentiel de ce qui est accessoire, les contraintes temporelles de la tâche, etc. Or l'expression d'"analyse par synthèse" tient pour acquis que le cerveau sache déjà ce qu'il fait, qu'il puisse comparer l'objet à comprendre et l'objet reconstruit : une fois une situation, sociale ou physique, interprétée de manière égocentrique, elle lui devient intelligible. Mais comment s'opère cette compréhension intra-cérébrale, et en quoi consiste-t-elle au juste ? Et comment le cerveau peut-il connaître directement ses

¹¹ Gordon (à paraître).

¹² D'autres théoriciens ont déjà analysé la contribution du système-miroir à la compréhension d'autrui. Cf. Gallese et Goldman (1998), Jeannerod (1999); Georgieff & Jeannerod (1998), Proust (2000), Jeannerod & Pacherie, (à paraître).

propres états (par introspection ??). Tel est le problème essentiel auquel est confronté le théoricien, quel que soit d'ailleurs le niveau où la simulation est censée s'opérer.

3.4. L'intermédiaire fonctionnel

Divers chercheurs, conscients du piège que nous tend la fausse évidence de la psychologie ordinaire, ont proposé une autre justification de l'impact de la simulation subpersonnelle sur la compréhension consciente de l'action ; ils font l'hypothèse qu'un niveau fonctionnel supplémentaire effectue la médiation entre le niveau subpersonnel de l'implémentation neuronale et le niveau personnel de la perception consciente et de l'action intentionnelle. Cette hypothèse a l'effet métathéorique souhaitable de prévenir les projections simplistes du niveau subpersonnel au niveau personnel. Ce type d'hypothèse "architecturale" a été développé à l'occasion des études de neuroimagerie rapportées dans la section 3.2., études qui manifestent, on l'a vu, un important recouvrement des régions cérébrales impliquées dans la génération de l'action, l'imagination de l'action et la perception des actions accomplies par d'autres agents. Ces résultats paraissent corroborer l'idée qu'il existe une forme d'équivalence fonctionnelle et structurale entre l'observation, l'exécution, l'imagination, l'imitation, et la planification de l'action. Marc Jeannerod, Jean Decety et leurs collaborateurs ont avancé qu'un processus simulatoire "couvert" — c'est-à-dire implicite et subpersonnel —, est activé dans toutes ces formes de traitement de l'information concernant l'action.¹³ Si la modélisation dynamique d'une action dans le format de son effectuation possible sert ainsi à effectuer des tâches très différentes, et à un niveau de traitement global de l'information (par opposition à l'activation des neurones-miroirs individuels), nous avons un début d'explication de la manière dont le cerveau extrait les informations dynamiques pour prédire les régularités comportementales. Ce que nous pouvons dire, c'est que les mêmes modèles prédictifs sont utilisés par le cerveau pour observer en vue d'imiter, pour exécuter et pour planifier l'action. L'hypothèse fonctionnelle a en outre le mérite d'être testable : on peut supposer a priori qu'une action qui ne peut être imitée — parce qu'elle est en dehors du répertoire du sujet, parce qu'elle n'a aucun sens pour lui ou parce qu'elle est biologiquement impossible — sera représentée différemment d'une action familière (soit qu'elle n'active pas le réseau de la simulation implicite, soit qu'elle active en outre des structures inhibitrices; cf. Decety & al., 1997 ; Grèzes et al., 1998).

La postulation de ce niveau fonctionnel "intermédiaire" fournit donc un outil d'analyse des étapes du traitement de l'information relative à l'action. Mais cette hypothèse soulève à son tour trois questions fondamentales: 1) le processus hypothétique de simulation implicite est-il fonctionnellement unique ou est-il décliné de manière différente selon le niveau de contrôle où il s'inscrit ? 2) En quoi permet-il d'offrir une justification du rôle d'un phénomène subpersonnel, la simulation couverte, dans la compréhension consciente de l'action ? 3) Quel rapport ce type de "simulation" a-t-il avec l'imagination personnelle ? Tels sont les problèmes théoriques auxquels il convient d'apporter une solution si nous voulons mieux comprendre comment s'opère la transition du subpersonnel au personnel.

¹³ C'est à une déficience de ce processus subpersonnel de simulation que seraient dûs les troubles de la conscience de l'agir chez le schizophrène (en particulier l'impression d'avoir ses propres actions contrôlées de l'extérieur). cf. Daprati et al., 1997.

4. La simulation: le problème de l'unicité

4.1. N'y a-t-il qu'une forme de simulation implicite ?

La question de l'unicité des processus simulatoires subpersonnels peut sembler être une question étrangère au philosophe, mais la réponse apportée contraint très largement les deux autres questions. En effet, si la simulation est réalisée par des moyens multiples et hétérogènes, on pourra difficilement supposer que le sujet puise sa connaissance de soi et d'autrui à une source unique ; on sera moins porté à estimer qu'un "modèle de soi" est à l'origine de toute forme de compréhension mentalisatrice.¹⁴ La trajectoire du subpersonnel au personnel devenant tortueuse et indirecte, il deviendra d'autant plus important de manifester la dépendance des jugements personnels à l'égard de la diversité des sources informationnelles et des facteurs motivationnels.

L'ubiquité de la simulation au niveau subpersonnel résulte des contraintes que doivent nécessairement respecter les systèmes visant à contrôler l'environnement et/ou à s'auto-réguler. Contrôle et régulation sont les fonctions grâce auxquelles le système régulateur peut atteindre un ensemble de buts dans l'environnement régulé tout en restant dans son domaine de viabilité (sans consommer trop de ressources, se mettre en danger, ou exiger un temps de traitement inadapté). Les théories du contrôle distinguent deux grands modes de régulation : le contrôle rétroactif et le contrôle direct ou prédictif. Le contrôle rétroactif procède par une correction successive des erreurs fondée sur la comparaison de la valeur réelle (feedback) de la propriété contrôlée par un système et de la valeur désirée. Pour que le contrôle rétroactif puisse être efficace, il faut notamment que le feedback soit fiable et disponible dans des délais brefs afin de permettre une correction rapide des déviations. Si tel n'est pas le cas ce mode de contrôle donne lieu à des phénomènes d'oscillation et d'instabilité. Dans le contrôle prédictif, le système de contrôle dispose d'un modèle du système à contrôler qui lui permet d'anticiper les conséquences des actions de ce système. Le choix et la correction des actions dépend donc non d'une comparaison entre valeur désirée et feedback mais d'une comparaison entre valeur désirée et valeur prédite comme résultat des différentes actions possibles. Un tel mode de contrôle présente plusieurs avantages. Le contrôle peut être efficace même lorsque la transmission du feedback est lente bruitée, ou intermittente. Il est également utile dans des situations potentiellement dangereuses où les conséquences d'une action peuvent être évaluées avant que celle-ci soit éventuellement exécutée. Toutefois, à l'évidence, l'efficacité du contrôle prédictif dépend de la prévisibilité des conséquences des actions du système contrôlé. En situations d'imprévisibilité, le contrôle rétroactif peut se révéler plus efficace. La régulation qui s'est largement imposée dans le vivant, et en particulier dans l'activité cérébrale, est celle qui s'effectue par un contrôle direct prédictif du système qui cause le processus considéré (*cause-controlled regulation*) plutôt que par un contrôle purement rétroactif. Le régulateur optimal le plus simple est celui dont les actions sont en correspondance avec les actions du système à réguler (Conant & Ashby, 1970). Ainsi, ce régulateur doit être capable de simuler de manière interne les interactions qu'il projette d'avoir avec son environnement. Chaque fois que le cerveau doit réguler une entité qui fonctionne comme lui-même (un autre cerveau, c'est-à-dire un autre homme ou animal), ou bien un sous-système du cerveau), il procède ainsi à la simulation dynamique de l'activité ou du domaine à réguler. Comme on le verra plus bas, il parvient également à simuler des entités dont la dynamique n'a rien de biologique.

¹⁴ Cf. Baars, 1988, Vogeley & Newen, 2003, Metzinger, 2003.

Dans le domaine de l'action, la régulation par simulation joue un rôle évidemment central; elle se manifeste dans les recouvrements fonctionnels évoqués plus haut. Le système moteur doit pouvoir engendrer le mouvement efficace dans un environnement changeant donné; il lui faut pour cela produire une simulation des commandes et de ses conséquences sensorielles (par un modèle forward ou prédictif); il doit en parallèle calculer les modèles inverses qui - étant donné l'état présent du système et de l'environnement et l'état désiré - permettent l'estimation des commandes motrices qui conduisent à obtenir ce dernier. La comparaison effectuée par simulation du modèle forward entre les réafférences souhaitées et celles qui sont engendrées par une commande donnée permet d'améliorer l'action correspondante. Une fois mis en place, ce mécanisme général d'évaluation peut être mis en œuvre pour apprécier l'efficacité du comportement d'autrui, pour imaginer mentalement que l'on agit ou pour apprendre une action à imiter.

Notons que ces simulations sont exécutées sans que le sujet en prenne conscience. Le sujet peut parfois pourtant avoir l'impression d'hésiter avant d'agir; se préparer, par exemple, à sauter un fossé assez large. La personne "se donne le temps de la réflexion", tandis que son cerveau est occupé à comparer l'efficacité de diverses commandes motrices dans les circonstances précises de l'action. Nous reviendrons plus bas sur la double description du même processus (aux limites entre subpersonnel et personnel).

Mais la simulation comprise comme le moment projectif d'une boucle de contrôle recouvre également beaucoup d'autres domaines d'activité cérébrale qui ne sont pas liés à l'action motrice. Elle est par exemple mise en jeu dans la métamémoire : le sujet qui doit retrouver un souvenir lance dynamiquement une simulation de l'activité de recherche afin de déterminer si ce souvenir "évoque" ou non quelque chose pour lui. La simulation doit également pouvoir s'appliquer à des paramètres non biologiques : à des propriétés physiques, spatio-temporelles, d'événements avec lesquels le corps du sujet n'a pas d'engagement direct immédiat (comme les phénomènes météorologiques, la prévision de l'usure d'un matériau, de la maturation d'un fruit) ou encore des dynamiques sociales (l'éducation programmée des enfants, les projets de carrière etc). Là encore, des modèles forward doivent anticiper les conséquences de divers paramètres qui engagent des forces extérieures ou des circonstances institutionnelles complexes. Les diverses boucles de contrôle nécessaires à la régulation des divers champs d'opération doivent être partiellement hiérarchisées, pour atteindre la cohérence locale et globale des systèmes de commande. On peut donc conclure qu'il n'existe pas une seule forme de simulation couverte, qui serait la simulation des actions.¹⁵

La réflexion sur la variété de ces usages implicites de la simulation conduit à mieux discerner deux types bien différents d'usage des modèles internes. Une partie d'entre eux utilisent le mode de fonctionnement endogène d'une structure pour prédire les états d'une entité externe. C'est en particulier le cas, comme on l'a vu, des supports simulateurs soi-même (en particulier des émotions et des intentions d'agir). Une autre partie doit mémoriser des dynamiques non biologiques, afin de prédire des phénomènes extérieurs physiques ou

¹⁵ Même dans ce dernier cas, la méta-analyse des divers travaux d'imagerie cérébrale fonctionnelle consacrée aux activations communes à diverses utilisations de l'information motrice, réalisée par Grèzes et Decety (2001), montre que la structure fine de la simulation couverte est d'une grande sensibilité à la tâche. En simplifiant beaucoup, il existe deux classes de manières différentes d'engager les représentations de l'action. La première rassemble l'exécution, l'imagination mentale et l'observation de l'action (SMA, Cortex prémoteur dorsal, gyrus supramarginal et lobe pariétal supérieur). Ces diverses opérations mettent toutes en jeu la sélection d'un programme moteur en fonction d'un but. La seconde rassemble l'imagination mentale et l'observation de l'action accompagnée de l'intention d'agir (pré-SMA et gyrus frontal dorsolatéral).

sociaux, avec lesquels le corps propre n'a pas d'engagement direct. Dès lors, la question se pose de savoir si ces deux modes de simulation coopèrent pour permettre le développement de la cognition sociale.

4.2. Simulation couverte, contrôle et compréhension consciente de l'action d'autrui : justification du recours au subpersonnel.

La mise en évidence de structures fonctionnelles intermédiaires, inaccessibles à la conscience, permet non de résoudre les questions philosophiquement cruciales pour la compréhension mentalisatrice d'autrui, mais de réorienter la recherche : la reconnaissance de l'existence des autres esprits ne dépend pas seulement d'un raisonnement théorique par analogie puisqu'elle est en permanence déjà présupposée par les structures cérébrales "résonantes". L'attribution à autrui d'intentions n'a plus à s'opérer à partir de l'identification du mouvement proximal puisque la capture de l'information et la mémorisation ont pour objet l'action et ses conséquences distales. Notons que la recherche de ce en quoi consiste la connaissance d'autrui (ou de soi) ne se résume pas à citer l'existence d'un processus subpersonnel arbitrairement tenu pour déterminant : on ne passe pas directement de l'existence d'une structure à l'invocation d'une distinction philosophiquement pertinente. On ne prête pas non plus au mécanisme subpersonnel une interprétation homunculaire. Par exemple, on s'abstient de conclure que le cerveau "comprend mieux ce qu'il fait lui-même". Mais on éclaire la nature de la boucle de contrôle qui rend possible l'agir, et on identifie mieux la composante du monitoring conscient dans cette boucle.

L'existence de boucles de contrôle explique la nature de la prise de conscience occasionnelle qu'a le sujet des mécanismes subpersonnels engagés dans l'effectuation d'une action ou dans un processus de pensée. Prenons le cas de l'imagerie consciente de l'action. Il s'agit d'une forme de représentation personnelle : le sujet peut délibérément s'y engager, en modifier le cours, enrichir le contexte représenté, choisir d'interrompre sa visualisation mentale, la cultiver s'il est sportif ou musicien, etc. L'imagerie constitue l'un de ces points de contact entre le niveau personnel et subpersonnel, dont nous avons établi plus haut l'existence. Comment comprendre l'imagerie ? C'est une forme de "monitoring" endogène qui constitue le retour, ou comme on dit la "réafférence", d'une commande motrice lancée "offline". L'imagerie a une valeur d'apprentissage parce qu'elle forme une partie constitutive du cycle efférence-réafférence d'une boucle de contrôle. Ce qui vaut de l'imagerie vaut aussi, plus généralement, de la perception : ce qui fait que la conscience sensorielle est source d'intelligibilité pour le sujet, c'est qu'elle lui permet de voir le monde non comme une diversité inorganisée, mais comme un contenu ordonné par les boucles de contrôle. Perception et action sont en relation de codépendance régulatrice.

Il devient possible, dès lors, de saisir les conditions de possibilité qui permettent à un sujet de prendre conscience des faits liés à une boucle de contrôle : comprendre que l'autre souffre, ou qu'il a l'intention de prendre un objet donné, suppose l'intégrité des mécanismes subpersonnels de simulation. Par exemple, le sujet doit percevoir les résultats de ses actions comme en rapport avec ses intentions, soit, en termes techniques, "*comme des réafférences*" pour avoir le sentiment d'être l'agent. Ou bien il doit ressentir les composantes affectives d'une situation imaginée pour pouvoir attribuer à autrui l'émotion correspondante. Si les mécanismes subpersonnels de contrôle sont perturbés, le sujet se voit dépourvu du moyen de ressentir la différence entre les conséquences attendues de l'action propre et les conséquences inattendues de l'action d'autrui ou entre une action qui provoque du plaisir et une action qui suscite de la douleur.

4.3. Quel rapport ce type de "simulation" a-t-il avec l'imagination personnelle?

La thèse que nous proposons peut être résumée de la manière suivante. Le mécanisme développemental et phylogénétique dit de « redescription » pourrait former l'élément crucial de la réponse à cette question, la simulation mentalisatrice constituant la redescription de la simulation de l'agir. La redescription entraîne normalement une démodularisation, c'est-à-dire le traitement conjoint d'informations de sources diverses. Pour ce qui nous concerne ici, les deux formes de simulation, biologique et nonbiologique, peuvent être intégrées du fait de la redescription. En particulier, la simulation de l'agir peut être enrichie par la représentation des dynamiques sociales (simulation non biologique).

Nous empruntons ici la notion de redescription aux travaux de Karmiloff-Smith (1992) et à sa théorie du développement cognitif. Karmiloff-Smith défend l'idée que le développement des compétences de l'enfant dans différents domaines passe par différents stades, chaque nouveau stade étant caractérisé par un changement des ressources représentationnelles et computationnelles utilisées au stade précédent, ce qu'elle appelle un processus de redescription représentationnelle. Son modèle (modèle RR) distingue quatre niveaux de représentation et re-représentation, qu'elle appelle Implicite (I), Explicite-1 (E1), Explicite-2 (E2) et Explicite-3 (E3). Le niveau de représentation Implicite possède les caractéristiques suivantes : (1) l'information est encodée sous forme procédurale ; (2) Les encodages procéduraux sont spécifiés séquentiellement ; (3) les représentations nouvelles sont stockées séparément ; et (4) les représentations implicites sont parenthésées (*bracketed*) et ainsi il n'est pas encore possible de former des liens à l'intérieur du domaine ou entre domaines. L'idée générale est que lorsqu'une action atteint son but dans une situation donnée, cette séquence est stockée comme un tout et peut ensuite être utilisée lorsque cette situation se présente à nouveau. En revanche, les constituants de la séquence ou ce qui fait que la séquence marche ne sont pas accessibles. L'information contenue dans la procédure reste implicite.

Le modèle RR postule l'existence d'un processus réitératif de redescription représentationnelle, donnant lieu aux représentations E1, E2 et E3. La transition du niveau I au niveau E1 est le résultat d'une redescription sous un nouveau format compressé des représentations procédurales encodées au niveau I. Ces redescriptions sont des abstractions dans un code de plus haut niveau et à la différence des représentations de niveau I ne sont pas parenthésées. Les connaissances implicitement contenues dans les procédures sont alors représentées explicitement et deviennent manipulables. Les composantes des procédures et leurs conséquences spécifiques peuvent être reconnues, et leurs relations à d'autres composantes représentées. Enfin une fois que la redescription a eu lieu, l'utilisation de contrefactuels devient possible.

Toutefois, Karmiloff-Smith souligne que si les représentations de niveau E1 sont explicites au sens où leurs composantes sont manipulables par le système, elles ne sont pas pour autant accessibles à la conscience ou verbalisables. Pour cela, il faut qu'intervienne un nouveau processus de redescription dans un code plus général. Le niveau E2 correspond pour Karmiloff-Smith au niveau de redescription où les représentations deviennent accessibles à la conscience sans être encore verbalisables, dans la mesure où elles utilisent un code représentationnel similaire à celui des représentations de niveau E1 dont elles sont la redescription. Le niveau E3 est celui où les représentations sont à la fois accessibles à la conscience et verbalisables. Enfin, il importe de souligner que ces quatre niveaux de représentations peuvent parfaitement coexister.

Les niveaux I et E1 correspondent à des formes subpersonnelles de représentation et de computation, alors que les niveaux E2 et E3 sont des niveaux personnels de représentations.

Malheureusement, les travaux de Karmiloff-Smith mettent surtout l'accent sur la transition du niveau I au niveau E1 et disent peu de chose du passage aux niveaux E2 et E3. Nous voudrions ici esquisser quelques spéculations sur les modalités de ce passage. Auparavant, il paraît toutefois intéressant de noter, comme l'a fait Grush (1995), que la transition du niveau I au niveau E peut-être interprétée comme le passage d'une forme très primitive de modèle prédictif, fondé sur des tables de correspondance associant un état actuel et une séquence d'action à ses conséquences sensorielles, à un modèle prédictif de forme beaucoup plus sophistiquée qui simule les articulations et la dynamique du système dont il a en charge le contrôle.

L'exemple de l'imagerie motrice consciente que nous avons évoqué tout à l'heure semble supposer, dans les termes du modèle RR, une redescription de niveau E2, soit un niveau d'accès conscient (et donc personnel) mais non langagier. Prenant appui sur les propositions de Karmiloff-Smith selon lesquelles chaque niveau de redescription fait intervenir un mode de codage plus abstrait et un format représentationnel plus général permettant l'établissement de liens représentationnels nouveaux au sein du domaine et entre domaines, nous pouvons risquer les spéculations suivantes. Les modèles prédictifs de niveau E1 assurant le contrôle des productions du système moteur font intervenir un format représentationnel particulier, sensori-moteur et utilisant des référentiels spécifiques. L'idée d'un tel format sensori-moteur est bien établie notamment dans le domaine de la vision, où de nombreux travaux montrent l'existence de deux modes très différents de traitement des informations visuelles, associés à des circuits nerveux différents, ce que Milner & Goodale (1995) appellent la « vision pour la perception » et la « vision pour l'action ». Comme le souligne Jacob (2003), d'un point de vue computationnel et représentationnel, la perception et la reconnaissance visuelle consciente des objets d'une part et le contrôle des actions dirigées sur des objets d'autre part exercent des contraintes très différentes sur le système visuel. Dans une tâche de perception visuelle où le but est d'établir un lien entre un percept et des connaissances déjà stockées en mémoire sur diverses catégories d'objets, l'objet doit pouvoir être identifié à des distances et sous des angles différents et dans des conditions d'éclairage variable, les traits durables de l'objet doivent être encodés afin de fournir au système de reconnaissance visuelle des indices sur la catégorie de l'objet perçu. Dans une tâche visuomotrice de préhension d'un objet, le système doit localiser une cible dans un référentiel égocentré et fournir sur l'orientation, la taille et la forme de l'objet à saisir des informations visuelles qui doivent pouvoir être très rapidement réactualisées en fonction des mouvements de l'agent. Ce traitement spécifique de l'information visuelle est ce que l'on appelle la transformation visuo-motrice. Il est vraisemblable que le mode de codage représentationnel utilisé dans les modèles de contrôle de l'action est de même type que celui de la transformation visuo-motrice.

Toutefois, nos interactions avec des objets ne dépendent pas toujours seulement de leur forme, taille et orientation ; elles sont aussi déterminées par leur fonction. Saisir un stylo pour le déplacer ne suppose pas qu'on l'ait identifié comme stylo. Les informations exploitées par le système de vision pour l'action sont suffisantes pour accomplir cette tâche. Si, en revanche nous voulons saisir le stylo pour écrire, nous devons l'avoir identifié comme stylo, le mode de saisie correcte dépendant non de la forme seule mais de la relation entre forme et fonction. Il faut, dans un cas comme celui-ci, que vision pour l'action et vision pour la perception coopèrent, que l'action soit contrôlée conjointement par des informations visuo-motrices et par des informations perceptives et les connaissances encyclopédiques qui leurs sont associées. La construction de modèles prédictifs pour des actions « fonctionnelles » suppose donc l'utilisation d'un format représentationnel plus abstrait où des informations sémantiques peuvent être représentées. On peut faire l'hypothèse que le processus de redescription qui nous fait ici passer de niveau E1 au niveau E2 implique le passage d'un mode de codage sensori-moteur à un mode de codage plus général, perceptivo-moteur.

Cette redescription représentationnelle présente les avantages décrits par Karmiloff-Smith et notamment la possibilité de représentation de liens interdomaines. Les connaissances représentées dans d'autres domaines sous un format perceptif similaire, par exemple certains aspects de notre connaissance de la physique naïve, deviennent ainsi accessibles et exploitables pour le contrôle de l'action. A ce niveau de redescription, deux formes de simulation sont ainsi susceptibles d'être intégrées : la simulation biologique du comportement du système moteur et la simulation non biologique des comportements des objets physiques et, plus généralement, tout mode de simulation non-biologique qui fait intervenir un format représentationnel de même type. En outre, dans la mesure où ce format représentationnel est aussi celui des représentations perceptives conscientes, l'imagerie motrice consciente devient possible par l'exploitation offline de ces modèles prédictifs. De la même manière, on peut supposer avec Karmiloff-Smith qu'avec le passage à des représentations de type E3, verbalisables, de nouveaux liens peuvent être représentés et exploités entre tous les domaines où ce niveau de redescription a été atteint et que les représentations de l'action peuvent être intégrées avec des représentations culturelles et sociales véhiculées dans le langage.¹⁶

Nous avons ici mis l'accent sur deux distinctions. Premièrement, nous avons proposé une distinction entre 1) les formes biologiques de simulation qui font intervenir des modèles internes des structures endogènes à un système et exploitent ceux-ci pour prédire le comportement d'un système doté de structures similaires et 2) des formes non-biologiques de simulation où le système construit un modèle de la dynamique d'un système de type différent, physique ou social, pour être en mesure d'en prévoir le comportement. Deuxièmement, nous avons fait usage de la théorie de la redescription représentationnelle de Karmiloff-Smith pour distinguer différents niveaux, implicites et explicites, subpersonnels et personnels, de simulation. Nous avons aussi essayé de montrer qu'au fur et à mesure que l'on s'élève dans cette hiérarchie, la distinction entre simulation biologique et simulation non-biologique perd de son acuité dans la mesure où la redescription permet l'élaboration de modèles plus abstraits qui intègrent les connaissances contenues dans les modèles biologiques et non-biologiques de niveau inférieur et sont ainsi de nature hybride.

Ces distinctions nous paraissent pouvoir apporter un éclairage original sur le débat entre théoriciens de la simulation et théoriciens de la théorie. Certains ont vu dans ce débat une dispute essentiellement verbale. Tel serait effectivement le cas si l'on tient pour simulation tout usage de modèles internes à des fins d'explication et de prédiction du comportement et que l'on considère simultanément que tout modèle interne constitue une théorie du domaine qu'il modélise. La distinction entre simulation biologique et non-biologique permet de donner une dimension plus substantielle au débat, le théoricien de la simulation affirmant alors que notre compréhension des états mentaux d'autrui s'appuie sur une forme de simulation biologique, tandis que, pour le partisan de la théorie de la théorie, elle fait intervenir un modèle interne non-biologique appliquant les concepts de la psychologie ordinaire. Toutefois, dans la mesure où il existe une hiérarchie de modèles internes, c'est-à-dire de systèmes de contrôle, chaque niveau possédant des caractéristiques représentationnelles et computationnelles distinctives et entretenant des relations complexes avec les modèles de niveau inférieur, il est douteux que la réponse apportée au débat à un niveau donné soit généralisable aux autres niveaux. Ainsi, l'existence de processus simulatoires biologiques à des niveaux subpersonnels n'est nullement la garantie que la

¹⁶ Ce rôle fondamental du langage dans la redescription est également au cœur d'une nouvelle théorie qui défend l'existence de deux systèmes de mémoire de travail, le second, plus récent, recodant les résultats du système ancien, avec les avantages de la démodularisation qui s'ensuivent. Cf. Gruber, 2003.

mentalisation au niveau personnel relève nécessairement de mécanismes simulateurs de même type. On doit donc tenir compte de cette distinction de niveaux pour évaluer la portée des arguments de l'un et l'autre camp. Enfin, comme on l'a souligné, la redescription n'implique nullement la disparition du modèle inférieur redécrit. Il est donc tout à fait possible que selon les circonstances la compréhension d'autrui fasse usage de modèles de différents niveaux et types.

5. Conclusion

Ce chapitre avait pour objectif de s'interroger sur la pertinence réciproque des recherches menées en philosophie de l'esprit et en neurosciences. La distinction entre niveaux personnel et subpersonnel a été mobilisée à la fois par les autonomistes pour opposer les modes d'intelligibilité philosophique et neuroscientifique, et par les compatibilistes, pour montrer la complémentarité des approches tout en délimitant leurs sphères respectives d'investigation.

Pour défendre une distinction tranchée entre les niveaux, on a avancé que le niveau personnel a pour marque distinctive le rôle qu'y jouent soit la rationalité entendue au sens normatif, soit les considérations sémantiques. Or la pertinence de ces deux caractérisations est contestée, on l'a vu, par de nombreux travaux en sciences cognitives. Le critère d'accessibilité à la conscience qui prévaut aujourd'hui ne permet pas toujours d'obtenir une distinction nette entre le personnel et le subpersonnel. La frontière est labile et dynamique. De surcroît, les spécialistes des neurosciences cognitives refusent aujourd'hui qu'on les cantonne à l'étude des processus et états subpersonnels. Il en résulte une situation nouvelle pour le philosophe de l'esprit. Il ne peut pas accepter le séparatisme qui ignore les déterminants subpersonnels de la conscience ; il ne peut pas non plus se satisfaire de l'idée que les mécanismes subpersonnels se réduisent à implémenter les pensées et les actions conscientes. Du même coup, le philosophe paraît entrer en concurrence avec le spécialiste des neurosciences cognitives et être confronté à une redéfinition de son rôle.

L'analyse que nous avons proposée de la simulation était destinée à nous permettre d'aborder plus concrètement ce problème. A l'intersection de la philosophie de l'esprit et de l'action, des neurosciences, de la psychologie du développement et de la psychiatrie, la simulation renvoie à un ensemble de théories et d'expérimentations dans lequel les philosophes ont apporté des éléments importants de réflexion. Initialement conceptuelle, leur contribution a d'abord consisté à proposer diverses manières rivales de comprendre ce que recouvre le terme ; puis, à prendre la mesure des différences de niveaux. Progressivement, la tâche des philosophes s'est modifiée avec le progrès de la recherche en neurosciences, exigeant d'eux qu'ils revoient leurs propres définitions et effectuent de nouvelles distinctions auparavant inaperçues d'eux, mais que suscitent les données. La nature des rapports entre subpersonnel et personnel fait partie de ces problèmes proprement philosophiques. Ce problème s'éclaire, on l'a vu, quand on situe la simulation personnelle relativement aux divers mécanismes simulateurs tacites qui y contribuent, et qui d'ailleurs s'étendent bien au-delà.

Ce que nous avons suggéré, c'est que le lien vertical (du subpersonnel au personnel) qui intéresse le philosophe pourrait être modélisé en tirant les conséquences de deux faits. Le premier est qu'il existe un niveau intermédiaire entre l'organisation neuronale et l'organisation psychologique consciente, à savoir le niveau fonctionnel du contrôle hiérarchique. Le second concerne les mécanismes qui permettraient aux modules intervenant à la base de cette hiérarchie de se décroiser, et en particulier d'obtenir l'accès à des modules de représentation langagière et réflexive. La prise en compte de ces faits relance la réflexion philosophique sur de nouveaux thèmes, concernant, par exemple, l'unité de la conscience ou le rapport entre la perception et l'action. Le philosophe semble donc avoir

encore beaucoup à faire pour poursuivre son travail de clarification conceptuelle dans un domaine scientifique au développement duquel il n'est pas entièrement étranger.

Références

- Arbib, M.A. 2002. The mirror system, imitation, and the evolution of language. In. C. Nehaniv & K. Dautenhahn (eds), *Imitation in Animals and Artifacts*. Cambridge, Mass.: MIT Press.
- Arbib, M. A. and Rizzolatti, G. Neural expectations: a possible evolutionary path from manual skills to language. *Communication and Cognition*, 29: 393-424.
- Baars, B. 1988. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Bermúdez, J. L. 2000. Personal and subpersonal: A difference without a distinction. *Philosophical Explorations*, 2: 63-82.
- Buccino G., Binkofski F., Fink G.R., Fadiga L., Fogassi L., Gallese V., Seitz R.J., Zilles K., Rizzolatti G., Freund H.J. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience* 13: 400-404.
- Churchland, P.S. 1986. *Neurophilosophy, Toward a Unified Science of the Mind/Brain*, Cambridge, Mass., MIT Press ; trad. franç. sous la direction de M. Siksou, *Neurophilosophie, l'esprit-cerveau*, Paris, PUF, 1999.
- Conant R.C. & Ashby, W.R. 1970. Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1, 89-97.
- Davidson, D., 1980. *"Mental Events" Essays on Actions and Events*, Oxford: Oxford University Press, 207-227. Trad. fr. P. Engel, *Actions et Evénements*, Paris, Presses Universitaires de France, 1993, 277-304.
- Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J. C., & Fazio, F. 1994: Mapping Motor Representations with PET. *Nature*, 371, 600-602.
- Decety, J., J. Grezes, N. Costes, D. Perani, M. Jeannerod, E. Procyk, F. Grassi, and F. Fazio. 1997. Brain activity during observation of actions: influence of action content and subject's strategy. *Brain* 120:1763-1777.
- Dennett, D. 1969. *Content and Consciousness*. London: Routledge, Kegan and Paul.
- Dominey, P. Decety, J., Broussolle, E., Chazot, G. & Jeannerod, M. 1995. Motor imagery of a lateralized sequential task is asymmetrically slowed in hemi-Parkinson patients. *Neuropsychologia*, 33: 727-741.
- Fadiga, L. Fogassi, L. Pavesi, G. & Rizzolatti, G. 1995. Motor facilitation during action observation: A magnetic Stimulation study. *Journal of Neurophysiology*, 73: 2608-2611.
- Fogassi, L. & Gallese, V. 2002 The neural correlates of action understanding in non-human primates. In M. I. Stamenov & V. Gallese (eds), *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins, pp. 13-35.
- Gallese, V. & Goldman, A. I. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Neuroscience*. 2, 12: 493-501.
- Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. 1996. Action recognition in the premotor cortex, *Brain*, 119, 593-609.

- Georgieff N. & Jeannerod, M. 1998. Beyond Consciousness of external Reality. A "Who" system for consciousness of action and self-consciousness, *Consciousness and Cognition*, 7, 465-477.
- Grafton, S. T., Fadiga, L, Arbib, M. A., Rizzolatti, G. 1997. Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*, 6: 231-236.
- Grèzes, J. & Decety, J. 2001. Functional anatomy of execution, mental simulation, observation and verb generation of actions: a meta-analysis. *Human Brain Mapping*, 12: 1-19.
- Goldman, A. I. 1989. Interpretation psychologized, *Mind and Language*, 4, 161-85.
- Goldman, A. I. 1993. The Psychology of Folk Psychology, *Behavioral and Brain Sciences*, 16, 15-28.
- Goldman, A. I. 1995. Empathy, Mind and Morals, in Davies, M. & Stone, T., (dirs.), (1995), *Mental Simulation*, Oxford, Blackwell, 185-208.
- Gopnik, A. 1993) How we know minds, the illusion of first-person knowledge of intentionality. *Behavioural and Brain Sciences*, 16, 1-14.
- Gordon, R.M. 1986. Folk Psychology as simulation, *Mind and Language*, 1, 158-71.
- Gordon, R.M. 1996a. Simulation without introspection or inference from me to you, in Davies, M. & Stone, T., (dirs.), (1995), *Mental Simulation*, Oxford, Blackwell, 53-67.
- Gordon, R.M., 1996b. 'Radical' simulationism, in Carruthers, P., & Smith, P.K., (dirs.), *Theories of Theories of Mind*, Cambridge, Cambridge University Press, 11-21.
- Gordon, R.M. manuscrit. Intentional agents like myself.
- Grèzes, J., Costes, N. & Decety, J. 1998. Top-down effect of strategy on the perception of human biological motion : a PET investigation. *Cognitive Neuropsychology*, 15: 553-582.
- Gruber, O. 2002. The co-evolution of language and working memory capacity, in Stamenov & Gallese (eds.), *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins, 77-85.
- Grush, R. 1995. *Emulation and Cognition*. Doctoral Dissertation, University of California, San Diego.
- Hornsby, J. 1993. Agency and Causal Explanation, in J. Heil & A. Mele (eds)., *Mental Causation*, Oxford, Clarendon Press, 161-188.
- Hornsby, J. 1997. *Simple Mindedness*. Harvard: Harvard University Press.
- Jacob, P. 2003. Philosophie et neurosciences: le cas de la vision. Dans *La Philosophie Cognitive*, sous la direction de E. Pacherie et J. Proust. Gap: Ophrys, à paraître.
- Jeannerod, M. .1999,. To act or not to act, perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology*.52A :1-29.
- Jeannerod, M. et Pacherie, E. (submitted), Agency, simulation and self-identification.
- Karmiloff-Smith, A. 1992. *Beyond Modularity*. Cambridge, Mass.: MIT Press.
- Leslie, A. M. 1987. Pretence and representation, The origins of 'theory of mind'. *Psychological Review*, 94, 412-426.
- McDowell, J. 1985. Functionalism and anomalous Monism, republished in J. McDowell, (1998), *Mind, Value and Reality*, Cambridge, Harvard University Press, 325-340.

- McDowell, J. 1994. *Mind and World*, Cambridge, Harvard University Press.
- Metzinger, T. 2003. *Being no one, the self theory of subjectivity*, Cambridge: MIT Press.
- Milner, D. & Goodale, M. 1995. *The visual brain in action*. Oxford: Oxford University Press.
- Nagel, T. 1986. *The View from Nowhere*, Oxford, Oxford University Press. Trad. Fr. par S. Kronlund, *Le point de vue de nulle part*, Combas: Editions de L'Eclat, 1993.
- Nisbett, R. E. 2002. *The Geography of Thought: Why We Think the Way We Do?* New York: The Free Press.
- Peacocke, C. 1994. Content, computation and externalism. *Mind and Language*, 9, 3: 303-35.
- Perner, J., Baker, S. & Hutton, D. 1994. Prelief: the conceptual origins of belief and pretense. In C. Lewis & P. Mitchell (eds.), *Children's early understanding of mind*, 261-286). Hove, UK: Psychology Press.
- Perner, J., Gschaider, A., Schrofner, S., & Kühberger, A. 1999. Predicting others through simulation or by theory? A method to decide. *Mind & Language*, 14.
- Perner, J. & Kühberger, A. 2003 Putting philosophy to work by making simulation theory testable : the case of endowment. In Ch. Kanzian, J. QUITTERER, E. Runggaldier (eds), *Persons. An interdisciplinary Approach*, Wien: öbv-hpt. (Proceedings of the 25th International Wittgenstein Symposium, Kirchberg am Wechsel, Austria, 11 - 17 August, 2002.), pp. 153-168.
- Perrett, D. I. 1999. A cellular basis for reading minds from faces and actions. In M. Hauser & M. Konishi (eds), *Neural Mechanisms of communication*, Cambridge, Mass: MIT Press, pp.
- Proust J. 2000. "Awareness of Agency : Three Levels of Analysis", in T. Metzinger (ed.), *The Neural Correlates of Consciousness*, Cambridge: MIT Press, 307-324.
- Rizzolatti, G., Carmada, R., Gentilucci, M. Luppino, G & Matelli, M. 1988. Functional Organization of Area 6 in the Macaque Monkey. II Area F5 and the Control of Distal Movements. *Experimental Brain Research*, 71, 491-507.
- Rizzolatti, G., Craighero, L. & Fadiga, L. 2002. The mirror system in humans. In M. I. Stamenov & V. Gallese (eds), *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins, pp. 37-59.
- Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D. & Fazio, F. 1996. Localization of Grasp Representations in Humans by PET. 1. Observation versus Execution. *Experimental Brain Research*, 111, 246-252.
- Ryle, G. 1949. *The concept of Mind*, London, Hutchinson.
- Robinson, E.J. & Mitchell, P. 1995. Masking of children's early understanding of the representational mind: Backwards explanation versus prediction, *Child Development*, 66, 1022-1039.
- Sirigu, A., Duhamel, J. R., Cohen, L., Pillon, B., Dubois, B. & Agid, Y. 1996. The mental representation of hand movements after parietal cortex damage. *Science*, 23, 1564-1568.
- Stamenov, M. I. & Gallese, V. (eds). 2002. *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins.
- Stephan, K. M., Fink, G. R., Passingham, R. E., Silbersweig, D., Ceballos-Baumann, A. O., Frith, C. D., & Frackowiak, R. S. J. 1995: Functional Anatomy of the Mental Representation

of Upper Extremity Movements in Healthy Subjects. *Journal of Neurophysiology*, 73, 373-386.

Stich, S. 2003. Philosophie et psychologie cognitive. Dans *La Philosophie Cognitive*, sous la direction de E. Pacherie et J. Proust. Gap: Ophrys, à paraître.

Stich, S. 2001. Plato's Method Meets Cognitive Science. *Free Inquiry*, 21, 2: 36-38.

Vogeley, K. & Newen A. 2003. Mirror neurons and the self construct. In Stamenov & Gallese (eds.), *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins, 135-150.

Weinberg, J. Nichols, S. and Stich, S. 2001. Normativity and Epistemic Intuitions. *Philosophical Topics*, 29, 429-460.

Wimmer, H. and Perner, J. 1983. Beliefs about beliefs, Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.