

pascal engel

LA PSYCHOLOGIE COGNITIVE PEUT-ELLE SE RÉCLAMER DE LA PSYCHOLOGIE ORDINAIRE ?

« Quand nous parlons de la « psychologie comme science naturelle », nous ne devons pas supposer que cela veuille dire une sorte de psychologie fondée définitivement sur un sol ferme. Cela veut dire juste l'inverse ; cela veut dire une psychologie particulièrement fragile, et dans laquelle les eaux de la critique métaphysique fuient à tous les joints, une psychologie dont toutes les hypothèses de base et les données doivent être reconsidérées à partir de connexions plus larges et traduites en d'autres termes. C'est, en bref, l'expression d'un manque de confiance et non pas d'une arrogance... Une série de faits bruts ; un peu de bavardage et de spéculation sur des opinions ; un peu de classification et de généralisation au niveau seulement descriptif ; un préjugé fort que nous *avons* des états d'esprit, et que notre cerveau les conditionne... Ce n'est pas de la science, ce n'est que l'espoir d'une science... La meilleure manière dont nous puissions faciliter "une réalisation scientifique" est de comprendre combien grande est l'obscurité dans laquelle nous nous mouvons, et de ne jamais oublier que les hypothèses de science naturelle dont nous sommes partis sont des choses provisoires et révisables ».

William James, *Principles of Psychology*, 1982, p. 467, cité par Pylyshyn 1984, p. 272)

Les articles réunis ici traitent tous, sous un aspect ou sous un autre, des relations entre la psychologie ordinaire et la psychologie cognitive. Mon but, dans cette présentation, est d'essayer d'expliquer pourquoi le problème posé par la nature de ces relations, qui a été récemment au centre d'une vaste littérature, est à la fois un problème méthodologique central pour ce que l'on appelle les « sciences cognitives », et pour la philosophie de l'esprit et de la psychologie en général.

Quand nous expliquons, dans la vie de tous les jours, le comportement de nos semblables, nous utilisons toute une variété de termes psychologiques tels que « croire », « douter », « désirer », « craindre », etc. Nous supposons naturellement que ces termes dénotent des états mentaux internes aux sujets auxquels nous les attribuons et que ces états ont un certain contenu, que nous rapportons habituellement au moyen de phrases telles que « Il croit que Fidel est un sage », ou « Il désire que Khadafi se calme ». Suivant la terminologie

usuelle des philosophes, on peut appeler ces états mentaux des *attitudes propositionnelles*, parce qu'ils supposent une attitude du sujet par rapport à un contenu propositionnel (que *Fidel est un sage, que Khadafi se calme*). Nous supposons aussi que ces états d'attitudes propositionnelles sont, d'une manière ou d'une autre la *cause* du comportement et des actions que nous observons : il est allé vivre à Cuba *parce qu'il* croyait que Fidel est un sage, il a adressé un ultimatum à Khadafi *parce qu'il* désire que Khadafi se calme. La forme la plus courante de ces explications causales ordinaires implique la présence conjointe, chez l'agent, de croyances (ou tout au moins d'états par lesquels l'agent a une certaine information sur le monde ou suppose que tel état de choses est présent) et de désirs. Croyances et désirs interagissent pour causer les actions : il est allé à Cuba parce qu'il croyait que Fidel est un sage, *et* parce qu'il désirait aller vivre à Cuba. Ces états ont également un pouvoir prédicatif : étant donné cette croyance et ce désir, nous prédisons que l'agent ira vivre à Cuba. Bien sûr nous supposons aussi que plusieurs croyances ou désirs peuvent interagir, et qu'ils ne sont causes du comportement que si d'autres croyances ou désirs ne viennent pas entraver ceux que nous prêtons à l'agent : si par exemple, l'agent croit aussi que la vie à Cuba manque de confort, et si cette croyance est suffisante pour l'empêcher de désirer aller y vivre, nous ne prédirons pas que la croyance que Fidel est sage et le désir d'aller à Cuba sont suffisants pour conduire l'agent à aller vivre à Cuba. Ce schème commun d'explications est largement irréflecti et spontané : il fait partie de la conception intuitive que nous avons de nous mêmes, de nos actions, et de celles de nos semblables. On peut supposer, selon la terminologie maintenant usuelle chez les philosophes, que ce schème est au centre de la psychologie ordinaire ou psychologie du sens commun (*folk psychology*). Deux questions importantes se posent alors :

- 1) dans quelle mesure peut-on supposer que la psychologie ordinaire est une théorie ou un schème unifié d'explications et de principes et comment décrire ce schème ?
- 2) une psychologie scientifique évoluée peut-elle légitimement se prévaloir du même type d'explications et de principes que ceux qui sont supposés former la psychologie ordinaire ?

Supposons, pour l'instant, que la psychologie ordinaire soit bien un tel schème unifié. Jusqu'à la fin des années 1950, la psychologie scientifique du vingtième siècle a été dominée par le béhaviorisme. Selon les béhavioristes, la psychologie ordinaire est fautive : il n'y a pas d'états *mentaux* dénotés par les termes usuels d'attitudes propositionnelles, parce que ces états sont des dispositions (physiques) au comportement observable. Les gens ne croient pas ni ne désirent : ils agissent parce qu'ils ont été sujets à un certain renforcement des réponses produites par l'occurrence de stimuli répétés, et non pas parce qu'il y aurait des épisodes mentaux internes. Une psychologie scientifique béhavioriste est donc incompatible avec la psychologie ordinaire. Mais la plupart des psychologues béhavioristes ne se préoccupaient pas de savoir si leurs investigations avaient ou non quelque rapport avec la psychologie ordinaire : ils se contentaient de l'ignorer, écrivant leurs articles sur leur travail de laboratoire, avec leur vocabulaire comportemental. Skinner (1971) est une exception : il soutient que le béhaviorisme montre simplement que des termes tels que « libre », « digne », « responsable » ne s'ap-

pliquent pas aux humains, et donc que notre conception de nous-mêmes est radicalement fautive. Il adopte donc, vis-à-vis de la psychologie ordinaire, la position qui sera appelée *éliminativiste* : nous devons simplement nous débarrasser de notre vocabulaire psychologique mentaliste usuel. La position adoptée par un grand nombre de philosophes (comme Ryle), dans les années 50, est différente : selon eux, les termes mentalistes usuels peuvent être *redéfinis* avec, ou on peut montrer qu'ils ont en réalité la *même signification* que, les termes désignant ces dispositions au comportement. Par exemple « a peur » a le sens de « est disposé à adopter un comportement de fuite ». Ce « béhaviorisme logique » est une doctrine portant sur la *signification* des termes mentaux de la psychologie ordinaire, et cette doctrine soutient que, une fois les traductions « comportementales » opérées, la psychologie ordinaire redevient compatible avec une psychologie scientifique béhavioriste. Malheureusement, cette doctrine est fautive. Elle n'est pas fautive seulement parce qu'il y a de nombreux états mentaux (comme les sensations) qui ne semblent pouvoir être définis par aucun comportement observable spécifique, mais aussi parce que même lorsqu'une définition en termes comportementaux est possible, elle ne l'est que *moyennant* la postulation d'*autres* états mentaux, lesquels ne peuvent être eux-mêmes définis comportementalement que moyennant d'autres états mentaux, et ainsi de suite (cf. par exemple Davidson 1980).

Au moment même où les philosophes prenaient conscience de l'échec du béhaviorisme logique, la psychologie scientifique prenait ses distances avec le béhaviorisme en général. Une grande partie du comportement, humain en particulier, ne peut pas être analysée en termes comportementaux. Ce que l'on a appelé la « révolution cognitive » en psychologie, dans les années 1960 (mais qui avait commencé en fait bien avant), a consisté à réintroduire, au sein des théories expérimentales de la psychologie, les notions même que les béhavioristes tenaient comme inéluctablement suspectes parce que « mentalistes » : celles de *croyance*, de *connaissance*, de *plan*, de *but*, de *désir*, par exemple. Il suffit d'ouvrir un ouvrage de psychologie cognitive (e.g. Neisser 1967, ou Johnson-Laird-Wason 1977) pour le constater. La notion de « cognition » est évidemment difficile à définir, tous ceux qui s'en réclament ne veulent pas nécessairement dire la même chose quand ils l'emploient. Mais l'un des sens fondamentaux de cette notion est que les organismes entrent en contact avec leur environnement et ont certains comportements parce qu'ils sont capables d'avoir certaines informations, ou certaines connaissances, sur cet environnement, c'est-à-dire de se le représenter de diverses manières. L'un des postulats de base de la psychologie cognitive est donc qu'il y a certains états internes des organismes, qu'on peut traiter comme des *représentations mentales*, qui permettent à ces organismes d'emmagasiner une certaine *information* sur l'environnement, laquelle, combinée avec les plans ou les buts de ces organismes, sont la cause de leur comportement. Comme on va le voir, tout dépend de la manière dont on définira des notions telles que celles d'information, de représentation mentale, ou de but. Mais l'analogie avec la manière dont la psychologie ordinaire postule des états mentaux doués de contenu comme causes du comportement est frappante.

Elle est si frappante que de nombreux philosophes et psychologues cognitifs ont pu considérer que les explications en psychologie et en sciences cognitives n'étaient pas substantiellement différentes de celles que la psychologie du sens commun donne dans la vie de tous les jours quand il s'agit de se rendre compte des actions des agents. Selon ces philosophes et psychologues, la psychologie cognitive peut se prévaloir du même type d'explications intentionnelles du comportement humain que la psychologie ordinaire, en expliquant celui-ci au moyen d'attitudes propositionnelles qui ont les propriétés :

- i) d'être sémantiquement évaluables, c'est-à-dire d'avoir des contenus intentionnels qui ont des conditions de vérité ;
- ii) d'être douées de pouvoir causaux sur d'autres attitudes et sur le comportement ;
- iii) d'être telles que les généralisations implicites de la psychologie ordinaire en termes de croyances et de désirs soient largement vraies à leur sujet.

Fodor (1975, 1981, 1985, 1987) et Pylyshyn (1984, 1987) sont les principaux représentants de cette thèse. Selon eux, une explication en psychologie cognitive (et dans les autres sciences cognitives comme l'IA également) doit, comme les explications de la psychologie ordinaire, supposer que les agents agissent en vertu d'états intentionnels et de représentations doués de contenus, qui sont les causes de leurs actions. Les formes usuelles de ces explications disent par exemple qu'une créature fait *X* parce qu'elle a *inféré* que *P* sur la base de la *croyance* que *q*, et a fait *X* parce qu'elle avait effectué cette inférence et eu ces croyances, et parce qu'elle désirait (projetait, voulait, etc.) que *X* soit le cas. Pour qu'il y ait explication, il faut qu'il y ait au moins une loi ou une régularité légale. Selon Fodor et Pylyshyn, c'est parce qu'il y a des généralisations cognitives qui ont la forme : Si *X* croit que *p* et désire que *q*, alors, *mutatis mutandis*, *X* fait *A*, que tel comportement de type *A* peut être expliqué. Pylyshyn (1984) soutient que quand une explication de ce type peut être donnée pour un système traitant de l'information, ce traitement est « cognitivement pénétrable ». La pénétrabilité cognitive n'est autre que le critère de l'explication cognitive selon cette conception. C'est parce que les généralisations cognitives partagent cette forme de base avec celles du sens commun que l'on peut dire, selon Fodor, que la psychologie cognitive peut se réclamer du même type d'explications intentionnelles que le sens commun. Mais ceci n'est cependant que la moitié de l'histoire. L'autre partie n'a pas grand chose à voir avec le sens commun. Quand ce dernier fait ces généralisations, il ne se préoccupe guère de la nature des représentations impliquées par les états intentionnels qu'il postule, et il suppose seulement qu'il y a un lien causal *quelconque* entre états mentaux, et entre états mentaux et comportement. Au contraire la psychologie cognitive essaie de spécifier quels types de représentations sont en présence, et quel est leur pouvoir causal. Selon Fodor et Pylyshyn, les attitudes ne peuvent répondre aux conditions (i) — (iii) ci-dessus que si (1) l'on tient que les organismes ont certaines attitudes qui sont des relations à des représentations mentales et que ces relations sont des relations *fonctionnelles* ou *computationnelles*, et si (2) les processus mentaux qui sous-tendent ces représentations sont des suites causales d'inscription de ces représentations dans l'esprit. (1) correspond à la thèse

fonctionnaliste en philosophie de l'esprit, selon laquelle les états mentaux ont des rôles causaux-fonctionnels à l'intérieur d'un système d'entrées-sorties, comparables aux états internes d'un ordinateur qui instancie un programme (Fodor 1975, 1987, Putnam 1975). Or les ordinateurs peuvent aussi nous fournir le modèle de la manière dont des « états » internes de la machine peuvent entrer dans des relations causales, car

Les opérations de la machine consistent entièrement en des transformations de symboles.

En effectuant ces opérations, la machine est sensible seulement aux propriétés syntaxiques (ou formelles, ou structurales) des symboles.

Et les opérations de machine sur les symboles sont entièrement limitées à des altérations de leurs formes (Fodor 1987, p. 19).

L'ordinateur, postule Fodor, est donc la métaphore, ou le modèle, approprié pour concevoir la nature de l'esprit, de ses processus, de ses représentations, et de leurs pouvoirs causaux. Selon cette « *théorie représentationnelle du Mental* » (TRM), les représentations sont encodées, stockées et manipulées sous la forme de symboles internes, ou d'un langage, comparable au « langage-machine » des ordinateurs, et douées de pouvoirs causaux, *en vertu de leurs formes et de leur syntaxe*. Fodor postule que ce langage doit être inné, privé, propre à chaque type de créature, qu'il est un « langage de la pensée » (Fodor 1975, 1987). Je laisserai ici de côté cette dernière thèse, dont le sens intuitif suffira à notre propos. Fodor et Pylyshyn prétendent que ces hypothèses sont largement conformes à tout le courant « classique » de recherches en sciences cognitives, et que cette théorie *représentationnelle-computationnelle* de l'esprit est la théorie qui sous-tend ce que nous pouvons appeler le « cognitivisme classique ». Pylyshyn défend, par exemple, la thèse selon laquelle il y a un médium unique de représentations, composé de symboles propositionnels (de phrases d'un langage interne), qui suffit à expliquer toutes les représentations cognitives, y compris celles qui ne semblent pas, de prime abord, être propositionnelles, comme les représentations perceptuelles et les images mentales (Pylyshyn 1984, cf. Engel à paraître).

Il va de soi que l'on ne saurait prêter au sens commun des hypothèses aussi sophistiquées. Quand les tenants du cognitivisme classique prétendent se réclamer des généralisations explicatives « ordinaires », ce n'est pas de cette partie sophistiquée de la « théorie » commune qu'ils parlent, mais du *schème* général croyance/désir qui, selon eux, la sous-tend. Il y a donc, sur ce point déjà, une ambiguïté dans la thèse selon laquelle la psychologie cognitive « conserve » le sens commun. Il faut comprendre: elle conserve la forme, mais pas nécessairement les contenus spécifiques de ses explications, c'est-à-dire les hypothèses (i) — (iii). En effet, même à l'intérieur de ce cadre classique, il y a toutes sortes de problèmes qui font douter que le sens commun psychologique puisse y retrouver ses « inspirations » intuitives. En premier lieu, comme le montre A. Woodfield ici même, le sens commun sympathise à priori avec l'idée qu'il y a aurait non pas une seule, mais plusieurs formes de représentations mentales: des concepts, des percepts, des images, des représentations verbales ou proposi-

tionnelles, chacune opérant à un niveau distinct et avec des formats distincts. La thèse de Fodor-Pylyshyn selon laquelle un seul médium de représentation — propositionnel — sous-tendrait toutes les autres n'est donc pas en accord avec le sens commun. Woodfield montre en outre que les recherches récentes en psychologie de la perception tendraient plutôt à confirmer ce point de vue du sens commun. Ceci montre, d'ores et déjà, si l'on accepte ce point, qu'il y a plusieurs manières de se réclamer du sens commun, ou que la psychologie cognitive « classique » à la Fodor-Pylyshyn n'est pas nécessairement confirmée dans toutes ses hypothèses de base. En second lieu, l'une des notions centrales de la psychologie cognitive est que la cognition relève d'un niveau représentationnel non conscient, « infra-personnel » ou « subdoxastique », par lequel les informations sont emmagasinées et traitées. Les attitudes propositionnelles invoquées à ce niveau ne sont pas les attitudes ordinaires de croyance, désir, etc. Ce sont des attitudes « tacites », ou « implicites », et le type de « connaissance » qu'elles mettent en cause a été appelé classiquement, surtout depuis les travaux de Chomsky et de ses disciples, « connaissance tacite » ou « implicite ». Martin Davies examine ici le statut de ces connaissances et montre qu'elles ne peuvent pas être mises sur le même plan que les connaissances propositionnelles ordinaires, notamment la connaissance réflexive que les locuteurs ont des règles de leur langage. Ceci cependant, ne menace pas, à mon avis, la thèse de Fodor-Pylyshyn, car cette thèse, comme nous l'avons vu, n'implique pas que les généralisations cognitives sur, par exemple la connaissance tacite des règles linguistiques, soient *les mêmes* que celles du sens commun. Elles peuvent être « cognitivement pénétrables » tout en conservant la forme (i- — (iii)), même si le sens commun ne les reconnaissait pas. Plus problématique, du point de vue qui nous occupe, est le recours que fait Davies à la notion fodorienne de *modularité*. C'est une notion complexe, controversée (cf. Garfield 1987), mais qu'on peut caractériser en gros comme suit : selon la conception architecturale de l'esprit de Fodor (1983), il faut distinguer deux sortes de systèmes autonomes dans la cognition : les systèmes « périphériques » (*input systems*) comme les systèmes reliés à une modalité sensorielle : vision, audition) etc., et le système « central », correspondant aux processus de pensée et de contrôle. Les premiers sont caractérisés par un traitement isolé (« encapsulé ») de l'information, et sont en ce sens des *modules*, alors que le second se caractérise par une intégration des informations, un « holisme » des inférences et représentations. Davies suggère que la modularité pourrait être l'essence réelle, sous-jacente des états de connaissance tacite. Fodor (1983) suggère que le domaine *par excellence* de la cognition est celui des systèmes modulaires de traitement d'information, et il avance une « loi » (1983, p. 140) : plus un processus cognitif est global (i.e. plus il est central, et moins il est « modulaire »), moins on sait de choses à son sujet. Or par définition les systèmes centraux sont ceux qui produisent des attitudes propositionnelles au sens usuel, tels que croyances, désirs, pensées, etc. Il devrait s'ensuivre que le domaine *propre*, ou en tout cas le domaine le mieux connaissable de la cognition est celui de systèmes périphériques modulaires, et pas celui des attitudes propositionnelles. Étant donné que le schème explicatif de la psychologie du sens commun fait

exclusivement appel à ces dernières et pas aux premiers, il semblerait que la psychologie cognitive ne soit pas un bon modèle pour une science de la cognition. Ici encore, cette difficulté peut être résolue si l'on distingue deux sens de la notion d'« attitude propositionnelle » : le sens d'états conscients du système central, et le sens des états symboliques inconscients (ou tout au moins auxquels on n'a pas accès) qui entrent dans diverses relations computationnelles postulées par TRM et la conception computationnelle de Fodor. Si l'on fait cette distinction, les « attitudes » dont parle Fodor quand il dit que la psychologie cognitive est conforme aux conditions (i) — (iii) ne sont pas les mêmes que celles dont parle le sens commun. En toute rigueur, au sein des sciences cognitives, seule la théorie de la décision — si elle fait partie de ces sciences —, qui parle de croyances, de désirs et de préférences plus ou moins conscientes des agents, s'appuierait sur la psychologie des attitudes du sens commun. Ceci montre (comme le note S. Guttentplan), que le schème de la psychologie ordinaire dont se recommande Fodor est, à tout le moins, fortement idéalisé.

Bien que la position métathéorique du Fodor-Pylyshyn ait été suffisamment influente pour être souvent considérée comme la conception reçue, elle a été contestée. Elle a d'abord été contestée du point de vue métaphysique au nom du dualisme. H. Robinson montre ici que la conception « fonctionnaliste » des états mentaux ne parvient pas à rendre adéquatement compte des sensations, des « qualités secondes » et des états subjectifs. Dans la mesure où les notions de ces états phénoménologiques font largement partie de l'appareillage conceptuel de la psychologie du sens commun, on peut dire que le cognitivisme classique les laisse de côté, et qu'il rencontre là une difficulté très sérieuse. Ensuite, du côté plus proprement des psychologues, j'ai mentionné les arguments rapportés par Woodfield contre l'idée d'un système unique de représentations mentales, et nombre de psychologues admettent ces arguments (pour les images : Kosslyn 1980 ; pour les représentations logiques : Johnson-Laird 1983). Mais la forme la plus radicale de contestation est venue de certains philosophes qui ont mis en cause l'ensemble des postulats (I) — (III) sur lesquels elle repose, et le postulat général que la psychologie du sens commun fournit la forme générale des explications cognitives. Ces philosophes défendent tous, à des degrés divers ce qu'on a appelé la thèse « éliminativiste », d'inspiration matérialiste selon laquelle le vocabulaire mental et intentionnel de la psychologie ordinaire devait non pas être expliqué ou réduit en termes d'une théorie représentationnelle-computationnelle de la cognition, mais purement et simplement *éliminé*. En suivant la classification adoptée par R. Bogdan, on peut distinguer au moins trois formes d'éliminativisme.

La plus radicale est celle de P. Churchland (1981, 1984, 1988) et de P.S. Churchland (1986). Selon ces auteurs, la psychologie ordinaire est bien, conformément à ce que dit Fodor, une théorie empirique, réaliste et descriptive du mental. Mais elle est, en tant que telle, fautive. Elle est fautive parce que l'explication cognitive doit pouvoir être réduite ultimement à une explication neurophysiologique, dans laquelle, l'idée même d'état mental et de contenu d'un état mental n'a pas de sens. Les neurosciences remplaceront, tôt ou tard, la psychologie, et tout

le vocabulaire mental qu'elle utilise sera finalement remplacé par des histoires de neurones, de synapses, etc. qui rendront compte complètement de ce que nous appelons à tort des représentations mentales.

Stich (1983) défend une autre version d'éliminativisme. Il est prêt à admettre la conception fonctionnaliste du mental de Fodor, et l'individuation syntaxique des états mentaux comme inscriptions de symboles dans l'esprit. Mais il rejette le thèse (I) du caractère évaluable sémantiquement des croyances, désirs, et autres attitudes ordinaires. Étudiant les manières dont le sens commun attribue des contenus intentionnels, il montre que ces contenus sont, au mieux, vagues, indéterminés, et incapables de recevoir des conditions de vérité claires. Il y a en ce sens une non correspondance systématique entre les explications ordinaires du comportement et les explications cognitives. Comme on le voit, Stich ne fait en un sens que radicaliser les remarques faites ci-dessus, sur les décalages entre les deux types d'explications. Selon lui, ces décalages sont suffisamment importants pour qu'on ne puisse pas dire que la psychologie cognitive pourrait se réclamer du sens commun.

Dennett (1978, 1987) défend la version la plus faible de l'éliminativisme (il soutient d'ailleurs que ce n'est pas vraiment de l'éliminativisme, cf. 1987 et plus bas). Il n'admet pas que la psychologie ordinaire soit une théorie fautive du mental, parce qu'elle n'est pas une théorie, ou plus exactement pas une théorie empirique. C'est un schème, normatif et idéalisé (donc ni vrai ni faux), qui règle nos attributions usuelles d'états intentionnels quand nous interprétons le comportement de toute créature. Comme n'importe quelle forme d'interprétation du comportement, la psychologie cognitive peut se prévaloir de ce schème, mais à condition de savoir qu'il n'est qu'une « instance intentionnelle », heuristique et instrumentale, permettant d'expliquer les actions (que ce soient celles d'un humain, d'une machine ou d'un animal). Mais il n'est pas question de dire que les états mentaux postulés par le schème seraient réels : il est même probable, conjecture Dennett, qu'une psychologie évoluée ne parviendra pas à trouver chez les individus des états correspondant à ce que nous appelons couramment des croyances, des désirs, ou des souhaits. Au mieux une explication cognitive réduira ces états à des conjonctions d'états infra-personnels (on peut le présumer : tacites) qui n'auront plus rien à voir avec les croyances et attitudes ordinaires. Celles-ci gardent néanmoins le pouvoir de *rationalisation* qu'elles ont dans toute interprétation.

Une autre forme de contestation de la thèse fodorienne a vu le jour dans le camp philosophique, celle qui vient des partisans d'une conception *externaliste* de l'intentionnalité. Elle n'est pas abordée dans les articles qui suivent, mais j'en dirai brièvement quelque chose. Fodor, en cela en accord avec de nombreux psychologues, soutient que les explications cognitives doivent obéir à ce qu'il appelle le « principe du solipsisme méthodologique », selon lequel les seuls contenus auxquels doit faire appel une psychologie scientifique sont des contenus internes, individuels, des organismes, sans référence aux états de l'environnement externe au sujet (Fodor 1980). Le problème est que les attributions usuelles de contenus d'attitudes propositionnelles ne semblent pas toujours conformes à ce principe « individua-

liste » ou « cartésien » : bien souvent nos attributions font appel à des entités du monde extérieur aux sujets, et on ne peut évaluer leur contenu que si l'on suppose qu'ils font sémantiquement référence à ces entités. C'est le cas de nombre d'attributions impliquant l'usage de termes indexicaux, comme « Jean croit que ce liquide est de l'eau ». On ne peut ici réduire le démonstratif « ce » à une description de propriétés de l'eau internes à la psychologie individuelle de Jean. Putnam (1975) a bâti sur ce thème une série d'expériences de pensée tendant à montrer qu'il y avait, dans les attributions usuelles d'attitudes une ambiguïté systématique entre l'attribution d'états mentaux internes au sujet (conformes au solipsisme méthodologique), et les attributions d'états évaluables sémantiquement par la spécification de conditions de vérité externes. Burge (1979, 1986) a généralisé cette thèse à tous les types de pensée, en soutenant que tous les contenus intentionnels sans exception, dépendent, pour leur individuation, de l'environnement. Cela menace la conception fodorienne de l'explication cognitive, parce que non seulement la sémantique des contenus intentionnels ne serait pas, dans cette conception « externaliste », dépendante de leur syntaxe, et parce que la psychologie cognitive, si elle est encore possible, devrait faire systématiquement référence aux propriétés de l'environnement. Le sens commun serait « sauvé », mais pas sous la forme voulue par Fodor.

Comme on le voit, tous ces débats sur le lien ou l'absence de lien supposés entre psychologie ordinaire et sciences cognitives dépendent largement des conceptions respectives que l'on a de l'une ou de l'autre.

Il est donc temps de revenir à la première question ci-dessus (1) : la psychologie ordinaire est-elle, comme le supposent aussi bien les partisans de la théorie représentationnelle de l'esprit que ses adversaires éliminativistes, une théorie ? Y a-t-il un ensemble de lois (même *ceteris paribus*), plus ou moins permanentes, gouvernant notre conception intuitive et pré-théorique du mental, et figurant dans les explications usuelles du comportement ? On peut en douter, pour les raisons mêmes qu'invoquent les éliminativistes eux-mêmes : s'il y a de telles lois, elles seront formulables en termes de relations causales entre des contenus (comme « si X a soif, et s'il croit que ceci est de l'eau, X boira ceci »), mais si les contenus intentionnels sont, comme le soutient Stich, indéterminés, comment pourra-t-on formuler ces lois ? Il est contradictoire, de ce point de vue, d'admettre qu'il y a une théorie en présence, si elle n'est pas spécifiable. On peut douter, si la psychologie ordinaire est une théorie, qu'elle soit aussi unifiée, cohérente et intégrée que des théories dans d'autres domaines (en physique, en biologie). Comme le note Guttenplan, si elle est une théorie, c'est en un sens très lâche, qui autorise de nombreux changements dans les concepts de base, et non pas une théorie permanente, « stagnante », comme le prétend Churchland. En ce sens, comment peut-on dire que la psychologie ordinaire des Grecs était la même que la nôtre ? Qu'aurait dit Euripide si on lui avait expliqué que sa Médée était victime du « stress » ? Ceci montre au moins que si psychologie ordinaire il y a, elle évolue. Par exemple, nous sommes, au vingtième siècle, peut être plus prompts à expliquer les comportements en termes de « situations » et de « contextes historico-sociologiques » qu'une certaine psychologie ordinaire du passé qui faisait plus facilement appel à des traits de

caractère, ou à des propriétés naturelles et a-historiques des individus. Si la plupart des humains actuels sont capables de comprendre les réactions des personnages d'un dessin animé de Disney ou de Tex Avery, jusqu'à tel point peuvent-ils comprendre les états mentaux de personnages de Balzac ou de Stendhal, qui, autant qu'on sache, ne parlent pas des martiens ? Et n'est-il pas ridicule — sinon dans le contexte d'une arrogance scientifique que d'aucuns pourraient trouver typiquement américaine — de dire que notre psychologie évoluée cognitive pourrait faire mieux pour expliquer le comportement que celle de Sophocle ? Le moins que l'on puisse dire est que certains philosophes américains auraient intérêt à suivre un cours d'histoire des mentalités. Ce n'est pas que cette notion de « mentalité » ne pose pas des problèmes méthodologiques de taille dans ce genre de travaux ; mais il est certain que la naïveté préthéorique n'est pas là où l'on croit. Mais on peut douter que la psychologie ordinaire soit une théorie pour d'autres raisons. Comme le souligne Bogdan, ce que l'on appelle la psychologie ordinaire est avant tout un schème pratique d'explications, et pas exclusivement et peut-être pas du tout une théorie. Peut être (comme l'a suggéré A. Morton 1980) la psychologie ordinaire n'est-elle qu'une collection de recettes, ajustées sans cesse au coup par coup pour les besoins des tractations de la vie courante avec nos semblables, « individualiste » au sens où elle ne repose ni sur des lois, ni sur des concepts généraux. On a le sentiment que même si la prédiction avancée par P.M. & P.S. Churchland, selon laquelle le vocabulaire *théorique* de la psychologie ordinaire pourrait dans le futur « disparaître » au bénéfice du vocabulaire des neurosciences, cela ne signifierait pas la disparition de l'aptitude pratique à expliquer et prédire les comportements, sur laquelle ce vocabulaire semble s'appuyer. L'éliminativisme ferait, en ce sens, une véritable erreur de catégorie.

On pourrait étendre ces remarques indéfiniment. Cela montre que la notion de « psychologie ordinaire » est, comme le dit Guttenplan, un label extraordinairement ambigu, sur lequel il est très difficile de faire reposer un argument. Si quelque chose de ce genre existe sous une forme unifiée, c'est avant tout la tâche de la psychologie sociale, de l'anthropologie culturelle, historique, et peut être d'une discipline encore à venir qui serait l'anthropologie cognitive (cf. Sperber 1985). Tant que l'on ne dispose pas d'une conception cohérente de ce que pourrait être la psychologie du sens commun, on voit mal comment on pourrait soit s'en prévaloir dans l'explication cognitive, soit la rejeter, et tout le débat paraît vicié.

Il y a cependant moyen de ramener ce débat dans des limites raisonnables. Il faut pour cela, si l'on admet que la psychologie ordinaire est quelque chose comme une théorie, distinguer, comme le fait Guttenplan, entre d'une part les théories et concepts psychologiques particulières par lesquelles, à telle ou telle époque, dans telles ou telles circonstances, un individu ou un groupe peut être amené à expliquer des comportements (par exemple la théorie et les concepts psychanalytiques — sous ses formes vulgaires : inconscient, phantasmes, complexes, etc. —, qui s'est plus ou moins intégrée à nos conceptions courantes aujourd'hui), et d'autre part le noyau théorique des explications usuelles du comportement. Fodor a, à mon sens, raison de souligner que ce noyau repose sur l'ascription d'états mentaux intentionnels aux

individus, sous la forme de généralisations du type de celles qui font appel à des croyances, désirs, et autres attitudes. Il serait par exemple tout à fait extraordinaire de constater qu'une culture donnée refuse d'attribuer des états mentaux, et en lieu et place de ceux-ci énonce des « explications » behavioristes comme « Il avait une disposition fortement renforcée au comportement de faire mouvoir ses jambes face à des stimuli X » ou matérialiste comme « ses neurones XYZ ont flashé » pour décrire la fuite de quelqu'un devant un gros chien, ou une pensée tendre pour sa bien aimée. Il se peut que cette seconde manière de parler repose, comme le soutiennent les éliminativistes sur un mythe (même s'ils admettent que c'est un mythe utile), un peu comme Ryle (1949) dénonçait, au nom d'une forme de behaviorisme logique, le mythe du « fantôme dans la machine ». Mais il est assez douteux qu'on puisse trouver des cultures qui auraient accompli ce genre de réformes. La question n'est pas tant celle de savoir si le noyau intentionnel de la psychologie ordinaire — c'est-à-dire le recours aux attitudes pour expliquer le comportement — est vrai ou faux que celle de savoir si on peut s'en passer. De ce point de vue, comme nous l'avons déjà noté, la psychologie des psychologues est mixte : quelquefois elle reprend des notions qui sont monnaie courante, comme celles de croyance, de plan ou de désir et, quelquefois elle forge ses propres notions, comme celle de système modulaire ou de transducteur. Quand elle reprend les notions usuelles elle en change le sens, plus ou moins. Il en est sans doute ainsi de toute discipline scientifique. Ceci ne signifie ni une allégeance, ni un refus de la psychologie ordinaire, comme le dit bien Pylyshyn (1984, p. xx) :

« S'il est vrai que la psychologie ordinaire est une mixture de généralisations, alors une théorie cognitive développée contiendra sans doute une certaine quantité de terminologie de psychologie ordinaire, mais elle en laissera sans doute une bonne partie de côté, tout en ajoutant ses propres constructions. Je ne vois pas comment le développement de cette image d'ensemble pourrait être anticipé en détail, sauf peut-être en psycholinguistique et en vision. Je ne vois donc pas pourquoi on devrait entreprendre une défense, ou un déni (cf. Stich 1983) de la psychologie ordinaire en général. »

Cette conclusion modérée sera ma conclusion officielle sur le débat, et je suppose que ce serait celle en général du psychologue. Mais, en tant que philosophe, je voudrais aussi proposer une image (à peine) plus spéculative.

Les positions, évoquées ci-dessus, de Churchland et de Dennett montrent que l'on n'est pas obligé d'accepter la thèse de Fodor respectivement sur la causalité des états mentaux et sur la réalité des attitudes et des contenus intentionnels. Churchland, en matérialiste, et conformément à la conception usuelle de la causalité, soutient que seuls des événements physiques peuvent être causes. Fodor, on s'en souvient, résout le problème en supposant que les états mentaux sont des relations à des symboles *physiques* inscrits dans le langage d'un organisme. Mais il y a une autre forme de matérialisme possible que celle de Churchland et que celle de

Fodor. Elle a été défendue par Davidson (1980), et elle est analysée ici par D. Laurier. Davidson soutient que tout événement mental concret est identique à un événement physique neurologique concret, mais il dénie qu'il y ait des lois strictes (déterministes) reliant des *types* ou propriétés d'événements mentaux à des types ou propriétés d'événements physiques. Cette double thèse (la thèse de l'identité des *tokens* physiques et mentaux) et celle de l'anomalie du mental (l'impossibilité de lois psycho-physiques comme de lois du mental) est ce que Davidson appelle le *monisme anomal*. On a soutenu qu'elle impliquait une forme d'épiphénoménalisme, c'est-à-dire d'inefficacité du mental. D. Laurier montre que ce n'est pas le cas, et que les événements mentaux restent des causes dans cette conception (conformément à la thèse de Davidson 1980). Cette position est importante, car elle montre que l'on n'est pas obligé de souscrire à une forme forte, réductionniste de matérialisme, comme Churchland. Cette thèse est en apparence compatible avec l'éliminativisme, parce que les seuls événements qu'il y a sont physiques. Mais ce n'est pas un éliminativisme. Davidson refuse, en effet, l'idée fodorienne que les généralisations descriptives de la psychologie ordinaire soient légales (« nomologiques »). Mais elle ne les rend pas pour autant stériles, comme Churchland. En d'autres termes, du fait que les événements mentaux sont des événements physiques (particuliers) on ne peut conclure que la psychologie intentionnelle ordinaire soit réductible aux neurosciences, pas plus qu'on ne peut conclure qu'il y a des lois du mental comme tel. Selon Davidson le schème ou le noyau des explications intentionnelles du comportement est indispensable à toute interprétation du comportement. Il n'a pas un statut descriptif, mais normatif, en ce sens que toute interprétation et explication psychologique dans les « sciences du comportement » — ou les « sciences morales », pour parler comme jadis, doit présupposer un degré minimal de rationalité des créatures que l'on interprète (cf. Engel 1989). Ce degré minimal est exprimé par les attributions de désirs et de croyances, et leur caractère causal. En ce sens, Davidson s'accorde avec la position de Dennett : la psychologie ordinaire a le statut d'une norme de rationalité. On ne peut s'en passer, non seulement parce qu'elle serait « utile » y compris aux fins d'une théorisation psychologique plus évoluée, mais encore parce que son schème de concepts est une condition *a priori* de toute interprétation. Ceci n'implique pas qu'elle doive être « éliminée » ou « conservée ». Ceci implique seulement que les états et les contenus que la psychologie ordinaire décrit ne sont pas réels au sens où Fodor soutient qu'ils le sont. Or, comme on l'a vu, il est très difficile, contrairement à ce que semble soutenir Fodor, de prendre à la lettre comme décrivant les processus et représentations analysés par la psychologie cognitive les états intentionnels de la psychologie ordinaire. Nous avons sans doute besoin ici d'une théorie plus précise de la manière dont nous attribuons ordinairement des attitudes propositionnelles à nos semblables. Il n'est pas certain que les seules considérations normatives suffisent (Gordon 1986), mais il paraît douteux que nous puissions interpréter le comportement d'autrui sans qu'au moins certaines normes de rationalité interviennent, même si elles ne sont pas aussi fortes que Davidson ou Dennett peuvent le dire. C'est pourquoi la conception

esquissée ici paraît la mieux à même de résoudre les tensions, et de concilier certaines des intuitions contradictoires que nous avons quant au statut d'une science du mental (qui n'existe pas encore) et d'une psychologie ordinaire (qui n'existe peut-être pas).

RÉFÉRENCES

- Burge, T. 1979 « Individualism and the mental », *Midwest Studies in philosophy, IV*.
 Churchland, P.M. 1981 « Eliminative materialism and propositional attitudes », *Journal of Philosophy*
 1984 *Matter and consciousness*, MIT Press
 1988 « Folk psychology and cognitive science », *Proceedings of the Aristotelian Society*, Supp. Vol.
 Churchland, P.S. 1986 *Neurophilosophy*, MIT Press
 Davidson 1980 « Mental events » (1969), in *Essays on actions and events*, Oxford U. Press
 Dennett 1978 *Brainstorms*, MIT Press
 1987 *The intentional stance*, MIT Press
 Engel 1986 « L'anomalie du mental », *Critique*
 1988 « Questions vives en philosophie de l'esprit », *Préfaces*, 10
 (à paraître) *Images mentales et psychologie cognitive* »
 1989 *La norme du vrai*, Gallimard
 Fodor 1975 *The language of thought*, MIT Press
 1980 *Representations*, MIT Press
 1983 *The modularity of mind*, MIT Press, tr. fr. Minuit 1985
 1985 « Fodor's guide to mental representation ; the intelligent Auntie *vade mecum* »
 1987 *Psychosemantics*, MIT Press.
 Garfield 1987, ed, *Modularity in knowledge representation and language understanding*, MIT Press.
 Gordon 1986 « folk psychology as simulation », *Mind and Language*, 1, 2
 Horgan, T, & Woodward, J 1985 « folk psychology is here to stay », *Philosophical Review*
 Johnson-Laird, P, 1983 *Mental models*, Cambridge University Press
 Johnson-Laird, P. & Wason, P. eds 1977, *Thinking*, Readings in cognitive science, Cambridge university Press.
 Kosslyn, S. 1980 *Image and mind*, Harvard V. Press.
 Laurier, D. 1987 « L'individuation des états intentionnels et le statut de la psychologie populaire », *Recherches sur la psychologie et le langage*, 8, Grenoble.
 Morton, A, 1980 *Frames of mind*, Oxford university Press
 Neisser, U. 1967 *Cognitive Psychology*, New York
 Putnam, H. 1975 « The meaning of meaning », *Philosophical Papers, vol, 2*, Cambridge university Press
 Pylyshyn, Z. 1984 *Computation and cognition*, MIT Press
 Ryles, G. 1949 *The concept of mind*, Hutchinson, Londres, tr, fr, Payot 1976
 Sperber, D. 1985 « Anthropology and psychology: towards an epidemiology of representations », *Man* 20
 Stich, S, 1983 *From folk psychology to cognitive science* MIT Press