

Le « patternalisme » de Dennett

Pascal Engel

Volume 22, Number 2, Fall 1995

URI: <https://id.erudit.org/iderudit/027328ar>

DOI: <https://doi.org/10.7202/027328ar>

[See table of contents](#)

Publisher(s)

Société de philosophie du Québec

ISSN

0316-2923 (print)

1492-1391 (digital)

[Explore this journal](#)

Cite this article

Engel, P. (1995). Le « patternalisme » de Dennett. *Philosophiques*, 22(2), 197–212.
<https://doi.org/10.7202/027328ar>

Article abstract

This paper discusses Dennett's instrumentalist theory of intentionality, and his recent attempt to account for our realist intuitions about the mind by using the notion of "patterns" underlying our intentional ascriptions. But the notion is too weak and indeterminate to do the job that Dennett intends to confer to it. He systematically refrains from discussing the exact relationship between the intentional level and more physical levels. Davidson's notion of pattern is more realistic, but is still wanting, if our realistic intuitions are to be maintained.

A R T I C L E S

LE « PATTERNALISME » DE DENNETT *

par

Pascal Engel

RÉSUMÉ : *On expose et discute dans cet article la théorie instrumentaliste de l'intentionnalité de Dennett. Ce dernier a tenté récemment de montrer que cette théorie n'avait pas les conséquences antiréalistes qu'on lui prête habituellement, en recourant à la notion de « trames » (patterns) sous-tendant les attributions en « posture intentionnelle ». On montre cependant que cette notion ne permet pas d'accomplir le travail que Dennett entend lui faire accomplir. Elle est trop indéterminée pour satisfaire les intuitions réalistes, parce que Dennett refuse d'envisager systématiquement les relations entre les niveaux intentionnel et physique. Davidson a aussi employé une notion de « pattern » dans sa théorie de l'interprétation, qui semble plus satisfaisante, mais qui ne lève pas les difficultés du « patternalisme » en philosophie de l'esprit.*

ABSTRACT : *This paper discusses Dennett's instrumentalist theory of intentionality, and his recent attempt to account for our realist intuitions about the mind by using the notion of « patterns » underlying our intentional ascriptions. But the notion is too weak and indeterminate to do the job that Dennett intends to confer to it. He systematically refrains from discussing the exact relationship between the intentional level and more physical levels. Davidson's notion of pattern is more realistic, but is still wanting, if our realistic intuitions are to be maintained.*

* Ce texte provient d'une conférence à l'Université du Québec à Trois-Rivières en février 1994, et a été écrit alors que j'étais invité du Département de Philosophie de l'Université du Québec à Montréal au semestre d'hiver. Je remercie pour leurs remarques Claude Panaccio, Nicolas Kaufman, Robert Nadeau et les participants du séminaire sur Dennett de l'UQTR.

Dans « Real Patterns¹ », mais aussi dans un appendice à « True Believers² » Daniel Dennett a mis un peu d'eau dans le vin de ce que l'on a appelé souvent son instrumentalisme au sujet des états et des contenus intentionnels. Selon la thèse instrumentaliste, il n'y a pas, littéralement parlant, de croyances, de désirs ou d'autres états désignés par la « psychologie populaire » : ces états ne sont que des postulats, heuristiquement posés, destinés à prédire le comportement sous la « posture intentionnelle ». Dans « Real Patterns », Dennett tempère ce que cette thèse peut avoir de choquant, et défend ce qu'il appelle une forme « atténuée » de réalisme au sujet des états intentionnels. Pour ce faire, il développe l'idée que la réalité de ces états est celle de *trames* (*patterns*) ou de structures, diversement interprétables et en ce sens indéterminées, mais néanmoins existantes. Je vais appeler, à la suite de Rey³, cette thèse « patternalisme ». Est-elle cohérente ? Est-elle bien un réalisme comme Dennett le prétend ? Répond-elle aux doutes que suscitait initialement l'instrumentalisme ou l'antiréalisme de Dennett quant aux états intentionnels ? Bref peut-on être patternaliste en philosophie de l'esprit ? La majeure partie de mon analyse sera consacrée à l'examen de cette thèse, et à l'analyse du réalisme modéré de Dennett. Il y a, chez Dennett et en général, beaucoup d'implications de ces thèses, en particulier pour son analyse de l'évolution⁴, pour sa critique de la notion d'« intentionnalité intrinsèque », pour la faveur qu'il accorde en général à l'IA, et pour sa théorie de la conscience (mais je ne ferai que signaler ici ce dernier point).

I

Je rappellerai brièvement d'abord en quoi consiste l'« instrumentalisme » de Dennett au sujet des états mentaux, et à quelles difficultés il se heurte. Dennett soutient, depuis principalement « Intentional Systems⁵ » que nous attribuons des états intentionnels (principalement des croyances et des désirs) du point de vue de la « posture intentionnelle » en traitant une créature *comme si* elle avait ces états intentionnels. Ces attributions nous permettent d'expliquer et de prédire le comportement des systèmes qui manifestent divers traits « intentionnels », tels que des humains, des animaux, et des ordinateurs. Dennett défend l'idée que ces attributions sont irréductibles à des attributions faites du point de vue physique (*physical stance*) et du point de vue du dessein

-
1. D. Dennett, « Real Patterns », *Journal of Philosophy*, 1991.
 2. D. Dennett, *The Intentional Stance*, MIT Press, Cambridge Mass, 1987, trad. fr. P. Engel, *La stratégie de l'interprète*, Paris, Gallimard, 1991.
 3. Rey, G. « Dennett's Unrealistic Psychology », à paraître in *Philosophical Perspectives*, 1994.
 4. Dennett, *The Intentional Stance*, chap. 8, et les articles contenus dans le numéro de *Lekton*, D. Fisette, dir. *Daniel Dennett et les stratégies intentionnelles*, Université du Québec à Montréal, 1992.
 5. in D. Dennett, *Brainstorms*, MIT Press, Cambridge Mass., 1978, trad. fr. J. Khalifa, *Philosophie*, n° 1, 1984.

ou du plan (*design*). En d'autres termes, plus traditionnels, il défend l'utilité explicative et prédictive des causes finales, par rapport aux causes efficientes et formelles. Mais ces attributions sont faites du point de vue d'un interprète, ou d'un attributeur, et sont relatives à ce point de vue. Elles font *comme si* un système intentionnel avait des désirs et des croyances, mais elles ne se prononcent pas sur la réalité de ces états.

La question de savoir si [un système décrit du point de vue la posture intentionnelle] a *réellement* des désirs et des croyances n'a pas de réponse⁶.

À strictement, parler, ontologiquement parlant, il n'y a pas de choses telles que des croyances, des désirs, ou d'autres phénomènes intentionnels. Mais les idiomes intentionnels sont « pratiquement indispensables », et nous devrions voir ce que nous pouvons faire pour rendre compte de leur emploi dans ce que Quine appelle un « idiome essentiellement dramatique »... Dans tout usage du vocabulaire intentionnel, on ne doit pas reconnaître des faits bruts simplement, mais un élément d'interprétation⁷.

Et à diverses reprises, il compare les croyances à des entités telles que des *voix* ou des *fatigues*. Littéralement parlant, il n'y a pas de choses telles que des voix, des fatigues. Il n'y a que des sons d'une certaine hauteur, des longueurs d'ondes, correspondant à ce que *nous* appelons des voix, et il n'y a que des états physiques (des quantités de globules blancs dans le sang, des états musculaires, des états cérébraux) correspondant à ce que *nous* appelons être fatigué.

Ces déclarations peuvent se prêter à plusieurs sortes d'interprétations. Dennett soutient-il une forme stricte de *fictionnalisme* ou d'*irréalisme*, selon lequel nos attributions d'états intentionnels (nos comptes rendus d'attitudes propositionnelles) sont *littéralement fausses*, et doivent être remplacées par des descriptions en termes comportementaux, ou en termes purement physicalistes ? En ce cas sa position serait très proche de celle des éliminativistes en philosophie de l'esprit, comme les Churchland ou Stich. Mais il ne défend pas cette position, car il insiste au contraire sur l'utilité, et l'indispensabilité du point de vue intentionnel de la psychologie populaire. Soutient-il alors plutôt que ces attributions, et les comptes rendus d'attitudes propositionnelles, ne sont *ni vraies ni fausses*, mais néanmoins utiles comme instruments de calcul et de prédictions ? Sa thèse serait alors une forme d'*instrumentalisme*, très comparable à l'*instrumentalisme* en philosophie des sciences, tel qu'il a été défendu par, entre autres, Bellarmine, Berkeley, Duhem, ou Van Fraassen, sous diverses formes. En philosophie de l'esprit comme en philosophie des sciences, cette thèse se heurte à des difficultés bien connues. Comment expliquer le succès des théories scientifiques si l'on ne dit pas que ces théories sont *vraies* ? De même comment expliquer le succès de la psychologie populaire si l'on ne dit pas qu'elle est *vraie* ? Comment pourrions-nous, en tant qu'espèce

6. *Ibid.*, chap. 1.

7. Dennett, *The Intentional Stance*, trad. fr., p. 342.

avoir survécu si nos croyances, nos désirs, etc. ne sont pas des entités réelles, que nos attributions désigneraient réellement ? Mais Dennett nous dit qu'il ne défend pas cette position instrumentaliste, parce qu'elle ressemble encore trop au fictionnalisme⁸. La réalité des états intentionnels ne réside pas seulement dans nos attributions, dans l'« œil de l'interprète ». Elle ne réside pas non plus dans ce que Reichenbach⁹ appelait des *illata*, c'est-à-dire des êtres posés en tant qu'entités théoriques (par exemple des électrons). Dennett préfère dire que les croyances, désirs, etc. de la psychologie intentionnelle sont des *abstracta*, des objets abstraits, posés aux fins du calcul ou des constructions logiques¹⁰. Il donne comme exemple les centres de gravité, définissables en termes de forces physiques. Mais il faut faire, nous dit-il¹¹, une distinction entre des *abstracta* utiles explicativement, et d'autres qui ne le sont pas. Le centre de la population des USA, défini comme le point mathématique qui est à l'intersection de deux lignes telles qu'il y a autant d'habitants au nord qu'au sud de la latitude, et autant à l'est qu'à l'ouest de la longitude est un *abstractum* en ce sens, comme peut l'être le centre des chaussettes perdues par Dan Dennett, le point défini comme centre de la plus petite sphère inscrite autour de toutes les chaussettes qu'il a perdues dans sa vie. Les croyances et les désirs sont des *abstracta* comparables aux centres de gravité, car ils sont utiles pour la prédiction du comportement, tout comme les centres de gravité sont de « bons » objets abstraits en physique. Mais la distinction entre des *abstracta utiles* et ce que l'on pourrait appeler des *abstracta futiles* ne nous rapproche pas beaucoup du réalisme : car si l'on dit que le critère de démarcation est l'utilité explicative, scientifique, d'un concept abstrait ou théorique, cela veut-il dire que cette utilité se mesure en termes du pouvoir de prédiction et du succès des concepts, ou bien en termes de leur adéquation à une réalité indépendante ? Tout dépend de la conception que l'on a de l'explication scientifique : ce concept est-il essentiellement épistémique ou ontologique ? Qualifier les croyances d'*abstracta* au même titre que des centres de gravité ne nous dit rien tant qu'on n'a pas précisé ce point.

Dennett nous dit qu'il n'est pas un instrumentaliste, mais un réaliste car « la croyance est un phénomène parfaitement objectif ». Mais il ajoute dans la

8. On pourrait aussi comparer ces discussions à celles qui ont cours actuellement au sujet du réalisme et de l'antiréalisme des propriétés morales, la position irréaliste étant défendue par Mackie, et une forme de position comparable à celle de Dummett étant défendue par Blackburn, *Spreading the Word*, Oxford, Oxford University Press, 1984, sous le nom de « quasi réalisme » Cf. plus bas. On peut aussi penser au débat au sujet des entités mathématiques qui oppose le platonisme réaliste à l'instrumentalisme de Field (*Science without numbers*, Blackwell, Oxford, 1984).

9. Cité par Dennett, *The Intentional Stance*, trad. fr., note 2, p. 72.

10. Dans « three kinds of intentional Psychology », Dennett, *Ibid*, p. 77, Dennett dit quelque chose de distinct : il dit que les croyances sont « à mi-chemin entre les *illata* et les *abstracta* ».

11. D. Dennett, *ibid*, p. 28.

foulée que « l'on ne peut la discerner que du point de vue de celui qui se place du point de vue d'une certaine *stratégie prédictive* » et que « son existence ne peut être confirmée que si l'on se place du point de vue du succès de cette stratégie¹² ». Qu'est-ce que cela veut dire ? Comment la croyance ne peut-elle être confirmée qu'à partir du point de vue d'un interprète et en même temps être « un phénomène parfaitement objectif » ? La seconde déclaration va dans le sens d'un vérificationnisme, alors que la première va dans le sens d'un réalisme. Comment peut-on défendre les deux à la fois ?

Il est plus instructif de considérer la relation entre les *abstracta* et le type de réalité par rapport à laquelle ils sont abstraits. Dans le cas des centres de gravité, c'est clairement une réalité physique, celle de forces physiques réelles, bien que le centre de gravité ne coïncide avec aucune entité entrant dans la composition physique des objets, comme une particule ou une sous-particule. De même que les centres de gravité peuvent être reliés à des forces physiques, les croyances et les désirs peuvent être reliés à des structures physiques objectives. Mais lesquelles ? Dennett ne veut pas dire que les croyances sont des états *identiques* à des états physiques d'un organisme, sans quoi il défendrait une forme de théorie de l'identité esprit-cerveau. Dans de nombreux textes il semble plutôt se commettre à une forme de fonctionnalisme et de théorie de l'identité *token/token* : les événements mentaux particuliers sont identiques à des événements physiques, mais les propriétés mentales ne sont pas identiques à des propriétés physiques. Mais admet-il, l'existence de propriétés mentales, ou de propriétés fonctionnelles auxquelles s'identifieraient les croyances ? Ce que Dennett va appeler des *patterns* doit nous permettre, selon lui, de répondre à ces difficultés.

Dans « Real Patterns », il introduit cette notion au moyen de deux exemples. Le premier est celui d'une série de six cadres, composé de points. L'arrangement des points sur les six cadres est différent dans chaque cas. Mais ils produisent tous la même trame ou *pattern*, de cinq carrés. Il appelle cela un « code barre ». (Si l'on voulait prendre un autre exemple, on pourrait prendre celui-ci. Imaginez, par exemple, que votre imprimante ne marche pas bien, et qu'à six reprises, vous produisiez une impression de votre texte ; dans chaque cas, les taches sur le papier sont distinctes, et l'encre s'est répartie différemment — dans certains cas elle a foncé sur une surface, dans d'autres une autre, etc. — mais l'impression produit quelque chose qui, dans chaque cas est votre texte, c'est-à-dire les lettres que vous aviez sur l'écran de votre ordinateur quand vous avez écrit le texte électroniquement). Selon Dennett, dans ces cas, le *pattern* est réel, même dans les cas où il est indiscernable à l'œil nu. Il est indiscernable, mais toujours reconnaissable. Pourquoi ? Parce que nous avons une mesure précise de l'information contenue dans chaque assemblage de points : la « carte à bits », qui identifie chaque point numériquement. Il utilise la définition de l'aléatoire mathématique de Chaitin : une série de points est

12. D. Dennett, *ibid.*, p. 27.

aléatoire si et seulement si l'information requise pour transmettre la série est incompressible : rien de plus court que la carte à bits ne préservera la série. Une série n'est pas aléatoire s'il y a une manière plus efficace de la décrire. En ce cas elle a un pattern. Le but de l'analogie est le suivant. Différents individus peuvent percevoir le pattern différemment, identifier des patterns distincts. Mais le pattern est là, néanmoins, il a une existence indépendante des observateurs.

Le second exemple que donne Dennett est celui du jeu de la vie de Conway. Je ne peux ici le présenter que très rapidement. C'est un jeu où l'on a un damier, ou un échiquier, composé de cases, donc chacune est activée ou désactivée (marche/arrêt) à un moment donné. La seule règle du jeu est que chaque cas, pour déterminer quoi faire à l'instant suivant regarde combien de ses huit voisins sont activés : s'il y en a deux, la case reste dans son état présent, s'il y en a trois, elle s'active, et autrement elle se désactive. Cet univers est parfaitement déterministe. Il va y avoir diverses sortes de configurations que pourront prendre les cellules à travers l'évolution du jeu de la vie. Ce sont notamment les *gliders* et les *eaters* ; aucun d'eux ne s'identifie à un ensemble fixe de cases, mais à une trajectoire. Certaines configurations sont réidentifiables. Ce sont des patterns.

On pourrait donner d'autres exemples de patterns : les configurations que forment les vols d'oiseaux ou les essaims d'insectes, les taches sur un tissu, les régularités formées par les préférences alimentaires d'un groupe ou d'un individu, des formes dans les nuages, etc.¹³

Le but de l'analogie entre les attitudes propositionnelles de la psychologie populaire et des patterns est assez clair. D'une part il est destiné à infirmer une thèse que l'on prête naturellement à Dennett (ainsi qu'à Davidson, j'y reviens plus bas) : l'interprétationnisme, selon lequel il n'y aurait, dans nos attributions usuelles de croyances, de désirs, etc. *rien de plus* que ce qu'il y a dans les interprétations que nous en donnons. Selon cet interprétationnisme une ascription de croyance de la forme « X croit que p » aurait les conditions de vérité suivantes :

(a) « X croit que p » est vrai ssi un interprète de X, placé dans la posture intentionnelle, pouvait attribuer à X la croyance que p, de manière à prédire le comportement de X¹⁴.

13. Cf. P. Jacob, « Patterns, causalité mentale et intentionalité », *Les études philosophiques*, 1992.

14. Selon Fodor et Le Pore, *Holism, a Shopper's guide*, Blackwell, Oxford 1990, chap. 5 et B. Dahlbom, dir. *Dennett and his critics*, Blackwell, Oxford, 1993, chap. 4, l'interprétativisme contient aussi la thèse *projectiviste* selon laquelle « X croit que p » doit s'analyser en :

(b) « X croit que p » est vrai si X est dans l'état qui serait normalement la cause chez moi (l'interprète) de ma croyance que p.

Je ne crois pas que Dennett soit commis à (b) ni qu'il ait jamais défendu (b). Il s'y oppose au contraire dans les critiques qu'il adresse à Stich (D. Dennett, *The In-*

Selon cette conception, il ne semble pas nécessaire qu'il y ait, chez X, des états tels que des croyances et des désirs. Il semble suffire que l'interprète soit en mesure d'en attribuer, du point de vue de la posture intentionnelle. Dennett veut nier que sa théorie ait cette implication. Il soutient que si une trame, un pattern, discerné du point de vue de la posture intentionnelle est interprétable chez une créature quelconque, et s'il est possible de prédire le comportement de la créature à partir de ce point de vue, alors il doit y avoir quelque chose qui est là, un ordre réel dans les choses. (Dennett se réfère à la déclaration d'Anscombe quand elle dit que l'intérêt de l'analyse du raisonnement pratique et de l'action intentionnelle chez Aristote tient au fait qu'il « décrit un ordre qui est là chaque fois que des actions sont faites avec des intentions¹⁵ ». Ce n'est pas parce que nous interprétons avec succès qu'il n'y a rien que nous interprétons, que seules les entités postulées dans nos interprétations sont réelles. Il y a un ordre objectif, des régularités objectives que capturent les patterns intentionnels que nous imposons aux agents. Mais quel ordre ? « De quoi un pattern est-il le pattern¹⁶ ? »

II

Il ne peut pas être le pattern de quelque chose qui soit isomorphe au pattern que nous discernons. En d'autres termes, il ne peut pas être, dans les choses, le reflet du point de vue que nous adoptons pour discerner le pattern. Comme je l'ai déjà dit, cette position semble parfaitement compatible avec une certaine forme de fonctionnalisme, selon laquelle les croyances et les désirs s'identifient à leurs rôles causaux dans la production des actions, étant donné des *inputs* de l'environnement et d'autres états mentaux. Les fonctionnalistes admettent que ces états, même s'ils sont fonctionnellement définis (définis uniquement par leurs fonctions) sont cependant réalisés dans des structures physiques (du cerveau). Mais ils sont réalisés de manière « multiple », au sens où il n'y a pas de corrélation unique, un-un, entre les structures discernées au plan de la psychologie intentionnelle et les structures cérébrales. Cette idée de la « réalisabilité multiple » des états mentaux est explicitement défendue par Dennett dans des passages de *The Intentional Stance* comme le suivant, où Dennett imagine des martiens, omniscients sur les transactions qui se produisent dans le monde au niveau physique, mais incapables de discerner la structure intentionnelle et les régularités auxquelles elle correspond :

Prenez le cas où les martiens observent un agent de change, décidant de donner un ordre d'achat de 500 parts de General Motors. Ils prédisent les mouvements exacts de ses doigts quand il compose le numéro de téléphone, et les vibrations exactes de ses cordes vocales quand il émet son ordre. Mais si les Martiens ne voient pas

tentional Stance, trad. fr., chap. 6). Mais il est certainement commis à (a).

15. D. Dennett, « Real Patterns », p. 30.

16. *Ibid.*

qu'on aurait pu substituer aux objets particuliers en présence un nombre indéfini de structures différentes de mouvements de doigts et de vibrations de cordes vocales — même les mouvements d'un nombre indéfini d'individus différents — sans perturber les événements extérieurs du marché boursier, alors c'est qu'ils ont été incapables de déceler une structure véritable du monde qu'ils sont en train d'observer. Tout comme il y a une infinité de manières d'être une bougie — et on n'a pas compris en quoi consiste un moteur à explosion tant qu'on ne réalise pas que l'on peut visser sur ces douilles toute une variété de dispositifs différents sans affecter la performance du moteur — il y a de même une infinité de manières de donner des ordres d'achat de la General Motors, et il y a des variétés de dispositifs propres à chaque société qui sont autant de douilles dans lesquelles l'une de ces diverses manières produira à peu près le même effet que n'importe quelle autre. Il y a aussi des pivots spécifiques à chaque société, pour ainsi dire, où la direction que prennent les gens dépend du fait qu'ils croient que p, ou désirent que q, et ne dépend d'aucune façon de l'infinité des autres manières par lesquelles ils peuvent être semblables ou différents¹⁷.

Ici les martiens discernent les patterns objectifs de mouvements physiques, mais il sont incapables de discerner les patterns intentionnels. Ce que la fiction des martiens indique, c'est qu'il y a bien des patterns objectifs, même s'ils ne sont pas discernables du point de vue martien.

Mais un fonctionnaliste comme Fodor admet lui aussi la conception fonctionnaliste, la théorie de l'identité *token-token*, et l'idée de la réalisabilité multiple. Il admet lui aussi que la psychologie populaire nous fournit des patterns, qui correspondent à des patterns objectifs dans les structures physiques du cerveau. Qu'est-ce qui distingue alors sa position de celle de Dennett ? Ce qui distingue sa position de celle de Dennett est que les patterns sont pour lui discernables non pas seulement au niveau intentionnel et au niveau physique, mais aussi au niveau d'une syntaxe du langage de la pensée. Aux états que nous appelons des croyances correspondent non seulement, pour Fodor, des configurations causales, mais aussi des phrases du mentalais, identifiables par leurs propriétés formelles ou syntaxiques, qui correspondent aux contenus de nos croyances. Fodor pense donc qu'il existe, au plan de la syntaxe du mentalais, un ensemble de structures discernables qui correspondent aux structures des contenus discernés par la psychologie populaire. Il pense aussi que ces structures syntaxiques ont une sémantique elle-même discernable (que l'on peut spécifier en termes d'une relation quelconque de covariance avec des états de l'environnement, mais c'est une autre question). En d'autres termes, ce sont plus que des patterns : ce sont des structures parfaitement déterminées, à la fois au niveau « intentionnel » et au niveau « microscopique ». C'est précisément ce que nie Dennett. Il nie que les patterns de la psychologie intentionnelle correspondent à des structures syntaxiques déterminées dans le cerveau, et que ces structures aient une sémantique déterminée.

Il nie cela non seulement en raison des difficultés intrinsèques de l'hypothèse du langage de la pensée, mais aussi parce qu'il admet la thèse

17. D. Dennett, *The Intentional Stance*, trad. fr., p. 40-41.

quinienne de l'indétermination de la traduction. D'un point de vue sémantique, ou du point de vue de l'interprétation des contenus d'attitudes propositionnelles, deux ensembles distincts de prédictions en termes intentionnels (qui attribuent des contenus divergents) peuvent s'appliquer également à des comportements observés sans qu'il y ait de moyen de départager ces prédictions. Ils peuvent discerner deux structures différentes à partir d'une trame de comportement commun — par exemple que Raoul croit que les rutabagas sont des fruits et que Raoul croit que les rutabagas sont des légumes — sans qu'il y ait de *fact of the matter* quant à la structure ou au pattern qui correspondrait à ces attributions, c'est-à-dire qui permettrait objectivement de corréler les patterns à des trames objectives et donc de départager les deux attributions intentionnelles¹⁸.

À quoi servent, selon cette conception, les attitudes propositionnelles ? Elles sont, nous dit Dennett, reprenant une analogie de Churchland¹⁹ des instruments de mesure, des projections ou des transductions, dans l'idiome intentionnel des complexités des processus sous-jacents cérébraux qui sous-tendent les états que nous attribuons²⁰. Deux questions se posent. La première est celle de savoir si les projections et transductions en question ne pourraient pas, finalement, correspondre à des états déterminés dans le cerveau. Revenons à l'image des martiens. Dennett ne pourrait-il pas dire que, une fois que nous disposerions, comme les martiens, de tous les détails physiques (neurophysiologiques) nous serions en mesure d'opérer des corrélations terme à terme entre les patterns de la psychologie populaire et ceux qui existent objectivement dans le cerveau ? En ce cas la position de Dennett serait celle de Churchland, celle du supermartien qui, une fois qu'il a établi les corrélations — qu'il a en fait *réduit* un type de pattern intentionnel à un type de pattern physique — n'aurait plus besoin de la psychologie populaire. Il pourrait la jeter après usage. En ce cas Dennett serait un éliminativiste. Il y a certains passages dans lesquels il semble s'approcher de cette position, et celui sur les martiens pourrait aussi bien être lu en ce sens : bien que les martiens ne voient pas les choses du point de vue intentionnel, ils sont parfaitement capables de prédire tous les mouvements correspondant à l'achat d'actions de la General Motors, et en ce sens pourquoi pourraient-ils vraiment manquer quelque chose en n'ayant pas, comme les humains, la psychologie populaire ? Mais Dennett n'admet pas non plus la position de Churchland, non pas seulement parce qu'il pense que la psychologie populaire est beaucoup plus économique que les calculs physiques compliqués des martiens ou d'une science neurophysiologique idéale, et donc qu'elle est indispensable, mais aussi parce qu'il pense

18. D. Dennett « Real Patterns », p. 47, fait référence à l'analogie de Quine entre les formes éléphantiques de buissons et les attitudes propositionnelles.

19. « Eliminative materialism and propositional attitudes », *Journal of Philosophy*, 1979.

20. *Ibid.*, p. 45-46.

qu'il n'existe pas de point de vue objectif véritable à partir duquel on puisse faire une corrélation entre les patterns intentionnels et des patterns physiques. Selon Churchland, il doit y avoir un *fact of the matter* quant à la nature exacte de la projection des attitudes propositionnelles à partir des trames physiques. Pour Dennett il n'y en a pas.

La seconde question est : ne pourrait-il y avoir, malgré l'indétermination de l'interprétation radicale, une réduction de l'indétermination, en sorte que, malgré les divergences des instruments de mesure que constituent les attitudes propositionnelles, on puisse soutenir que deux schèmes d'interprétation divergents soient néanmoins des descriptions indirectes *de la même réalité*, c'est-à-dire des mêmes patterns objectifs ? Ce serait une position qui aboutirait à une conception assez similaire à celle de Churchland, le réductionnisme en moins. Davidson a lui aussi utilisé l'image des patterns et des instruments de mesure :

La situation [dans le cas des attitudes propositionnelles] est analogue à celle de la mesure du poids ou de la température...

Nous mesurons les autres esprits non pas (principalement) en assignant des nombres à leurs phrases, mais en faisant des assignations à ces phrases à partir d'un autre ensemble potentiellement infini, celui de nos propres phrases, le seul ensemble qui nous soit immédiatement accessible et qui ait la structure adéquate. Etant donné la richesse de cette structure, et la nature de ses connexions avec le monde, il n'est pas surprenant que s'il y a une façon d'assigner nos propres phrases aux phrases de quelqu'un d'autre il y en ait beaucoup d'autres, qui pourront représenter tout ce qui sera significativement pertinent. Nous pouvons dire, si nous le voulons, que l'interprétation ou la traduction est par conséquent indéterminée, ou qu'il n'y a pas de fait décisif (*fact of the matter*) qui permette de dire ce que quelqu'un signifie par ses mots. Dans la même veine, nous pourrions parler de l'indétermination du poids ou de la température. Mais nous accentuons normalement le positif en clarifiant ce qui est invariant à partir de notre assignation de nombres et en passant à l'assignation suivante, car c'est ce qui est invariant qui est empiriquement significatif. L'invariant est le fait décisif. Nous pouvons considérer la mesure du mental à la lumière de cette analogie.

Si cette analogie entre la mesure physique et l'interprétation est correcte, comme je pense qu'elle l'est, alors l'indétermination ne prouve pas par elle-même que les concepts mentaux (tout au moins les attitudes propositionnelles) soient différents des concepts physiques. Mais l'indétermination est directement liée à une différence réelle²¹.

Dennett n'accepte pas la position de Davidson, qui implique que malgré l'indétermination, deux patterns intentionnels ou mesures mentales équivalentes ne peuvent pas, en dernière instance, différer de manière substantielle quant à ce qu'elles visent ou décrivent, en sorte que pour Davidson, il y a bien un fait décisif plus profond quant à ce que nous attribuons à titre d'états intentionnel à des créatures. Dennett ne refuse pas de dire que nos attributions intentionnelles pourraient détecter, ou suivre à la trace, un ordre objectif. Mais il

21. Donald Davidson, « La mesure du mental », in P. Engel, dir., *Lire Davidson*, L'Éclat, Combas, 1994.

refuse de dire que cet ordre objectif pourrait être stable et déterminé, en sorte que l'on décrive ce que sont, en dernière instance, les croyances et les désirs réels des systèmes intentionnels. C'est pour cette raison qui assimile les patterns à des structures réelles, mais instables. Dans les patterns il y a toujours du bruit.

III

La conception dennettienne est-elle cohérente ? Peut-il nous convaincre qu'il existe une sorte de troisième voie entre le réalisme « industriel » de Fodor et l'éliminativisme adventiste de Churchland, qui serait constituée par son réalisme modéré des patterns ? Cela dépend de la cohérence même du concept de pattern et d'une question que je n'ai pas abordée jusqu'à présent, qui est celle de leur statut causal.

Revenons aux analogies initiales, avec les codes barres et le jeu de la vie. John Haugeland²² a remarqué qu'il y a deux oscillations caractéristiques propres au concept de pattern. La première est que quand Dennet parle de la réalité des patterns, il oscille entre deux thèses : a) que la réalité des patterns tient à la réalité de leurs éléments (les points sur le papier, les pixels sur un écran, les oiseaux ou les insectes dans un essaim), et b) que cette réalité tient à celle du pattern lui-même, de la régularité qu'il exemplifie. Il est clair que la réalité des éléments n'est pas en doute. Mais pris isolément, ce ne sont pas eux qui forment le pattern. Donc cette réalité est triviale, et ce n'est pas ce dont a besoin Dennett pour sa thèse. Ce qui a une réalité, c'est la configuration, ou plutôt ses *traits* ou ses caractéristiques. Mais ici aussi il y a deux hypothèses : a) ou bien ces traits sont des propriétés objectives du pattern, des exemplifications de propriétés réelles des configurations, ou bien b) ces traits sont des propriétés non pas du pattern lui-même, mais des propriétés de notre *reconnaissance* du pattern, des propriétés que nous discernons. Quand nous discernons une forme, par exemple un « V » dessiné par un essaim d'oiseaux dans le ciel, ou le profil d'un homme dans un nuage, ces propriétés n'appartiennent pas objectivement au pattern, mais sont des configurations que nous interprétons dans ce que nous observons. Il n'y a pas de propriété objective de faire un « V » pour l'essaim ou de dessiner le profil d'un homme pour les nuages. Dennett insiste fortement sur le fait que les patterns sont relatifs à ce que nous discernons en eux : c'est nous qui reconnaissons dans le code barre une série de cubes. Mais il ne fait pas de doute que si la réalité des patterns consistait seulement dans leur perception, ou dans les interprétations, ou dans les *voir comme* que nous discernons, les patterns ne seraient pas réels, et seraient seulement dans l'œil de l'observateur.

Mais je crois que cette objection n'en est pas une, du moins sous cette forme. Quand Dennett parle du caractère reconnaissable des patterns, il ne

22. S. Haugeland, « Pattern and being », in B. Dalbohm, dir., *op. cit.*

parle pas de leur reconnaissance effective par tel ou tel observateur : cette reconnaissance peut varier, vous pouvez au lieu d'un « V » fait par l'essaim d'oiseaux y voir une flèche, ou un angle, etc. Dennett parle plutôt de la reconnaissabilité *en principe*, dont le critère est le caractère non mathématiquement aléatoire. De ce point de vue, il y a une structure objective reconnaissable par « l'observateur maximal », au sens où l'ensemble de l'univers est un observateur potentiel²³. Et Dennett ajoute que la notion d'interprétation est sans doute trompeuse, car elle semble dire que la régularité constituée par le pattern est une régularité *mentale*. Mais ce n'est pas une régularité mentale. Il préfère parler de *transduction*, au sens où l'on en parle en électronique. Il y aurait ainsi reconnaissance d'un pattern parce qu'il y a une structure physique objective transduite dans le pattern observé par l'observateur, et ensuite susceptible d'être interprétée.

Fort bien. Mais cela pose quand même problème ; car l'idée de transduction suppose quelque chose comme une *loi* de la transduction : quand un certain courant électrique est transduit, par exemple, dans des sons dans un récepteur téléphonique, ou dans des signes sur l'écran de l'ordinateur, il y a une loi de la transduction, sans quoi on ne voit pas comment les choses se produiraient de manière à ce que la régularité physique soit associée à la régularité auditive, et des signes sur l'écran, puis à la régularité sémantique que nous leur conférons. Dans ce cas, comment Dennett peut-il éviter de dire qu'il y a des corrélations régulières entre des entités décrites au niveau mental et des régularités décrites au niveau physique ? N'est-il pas commis à soutenir une position qui pourrait se rapprocher de celle de Fodor (si l'on accepte aussi qu'il y a des régularités syntaxiques ou symboliques intermédiaires entre les régularités mentales et les régularités physiques) ou de celle de Churchland (qui se passe de ces intermédiaires) ?

Cette question est étroitement liée à celle du statut causal des patterns. Car d'une part si l'on admet, comme le fait Dennett, qu'il y a des régularités au niveau intentionnel et que ces régularités ont un pouvoir prédictif, doit-on conférer à ces régularités un statut causal ou pas ? Supposons qu'on leur en confère un. Alors la question qu'on peut poser d'autre part est : si on admet qu'il y a une forme de correspondance entre ces régularités intentionnelles et des régularités descriptibles au niveau physique, dans quelle mesure ces dernières ne vont-elles pas être considérées comme ayant une prééminence causale par rapport aux régularités décrites au niveau intentionnel ?

À la première question, Dennett répond clairement : oui. Il soutient certes que les *abstracta* de la psychologie populaire ne font pas partie de l'ordre physique, et qu'en ce sens ils n'ont pas le statut causal qu'on confère habituellement à des entités physiques. Mais il ne soutient pas que ces *abstracta* n'aient aucun pouvoir causal :

23. D. Dennett, in B. Dahlbom, *op. cit.*, p. 214.

Plusieurs interprètes du présent article ont supposé que la conclusion que je propose ici est que les croyances (ou leurs contenus) sont des épiphénomènes qui n'ont pas de pouvoirs causaux, mais c'est une erreur d'interprétation qui découle d'une conception trop simpliste de la causalité. Si l'on découvre un pattern prédictif du type de ceux que j'ai décrit on a *ipso facto* un pouvoir causal — une différence dans le monde qui fait une différence testable par des méthodes empiriques testables de manipulation variable. Considérez le pouvoir causal d'un panneau se lisant « Repas gratuit » placé dans la vitrine d'un restaurant et comparez son pouvoir dans un restaurant à New York et dans un restaurant à Tokyo. Le niveau intentionnel est de toute évidence le bon niveau où prédire et expliquer ce genre de pouvoirs causaux ; le panneau produit de manière plus fiable une croyance dans une population de croyants que dans une autre, et les variations dans la couleur ou la typographie du panneau ne sont pas aussi prédictives des pouvoirs d'attirer les foules que les variations dans la signification. Le fait que les régularités sur lesquelles ces prédictions réussies sont basées sont capturables de manière efficace (seulement) en termes intentionnels et le fait qu'elles ne soient pas dérivées de « lois de couverture » ne montre pas que les régularités ne sont pas causales. Cela montre seulement que les philosophes se sont appuyés sur des notions étroites de causalité dérivées d'une attention à quelques exemples tirés de la physique et de la chimie²⁴.

Admettons donc que les généralisations intentionnelles sont bien causales. Dennett souscrit explicitement à l'idée davidsonienne selon laquelle les raisons sont des causes. Il souscrit aussi à l'idée davidsonienne selon laquelle les raisons sont des causes sans que cela implique qu'il y ait des lois causales strictes (comparables aux lois physiques) de la psychologie populaire. Mais dans ce cas, pourquoi Dennett maintient-il aussi que les patterns intentionnels nous permettent de faire des prédictions fiables ? Comment peut-il y avoir un pouvoir prédictif sans des lois précises ? C'est difficile à comprendre, sauf si l'on admet que les prédictions de la psychologie populaire ne sont pas précises parce que les lois elles-mêmes ne sont pas précises, ou sont des lois *ceteris paribus*. C'est une conséquence que Davidson admet. Mais même si l'on concède ce point, et le précédent sur le caractère causal de ces lois, il reste deux problèmes de taille pour Dennett.

Le premier est celui posé par la correspondance entre les patterns intentionnels et les patterns objectifs physiques. Si l'on ne veut pas soutenir une position purement interprétationniste, comme on l'a vu, il faut admettre qu'il y a une corrélation, et une corrélation systématique, entre les deux types de patterns. Mais comment éviter, si l'on ne soutient pas que cette corrélation est une pure coïncidence, une coïncidence occasionnelle à la Malebranche, de dire que cette corrélation ne donne pas lieu à des lois, des lois psychophysiques, et comment ne pas réintroduire le spectre du réductionnisme qu'était supposé avoir écarté la notion même de pattern ? Et si l'on admet que cette correspondance ne donne pas lieu à des lois, mais qu'il y a deux séries causales qui ne se rejoignent pas en un point ou de points quelconques, comment

24. D. Dennett « Real Patterns », p. 43-44, note.

éviter ou bien la surdétermination (deux sortes de causes, les unes mentales les autres physiques concourant aux mêmes effets sur le comportement) ou bien une forme d'épiphénoménisme (si l'on soutient que l'un des types de causes — on peut le présumer les causes « intentionnelles » — a un statut plus faible ou plus étiole que l'autre — on peut le présumer les causes décrites au niveau des patterns physiques). Comment éviter le problème que Kim²⁵ appelle de « l'exclusion explicative » entre les deux niveaux de causalité ? Cette question ne se pose pas seulement pour le statut *explicatif* des causes intentionnelles ; elle se pose aussi pour le statut *ontologique* des entités entrant en relation causale, en l'occurrence les attitudes propositionnelles et leur contenu. Car si une entité a un statut causal, alors elle doit avoir un statut réel, elle ne doit pas seulement être une entité postulée *pour les fins de l'explication*, indépendamment du fait que les explications auraient des implications pour la réalité ontologique des processus qui entrent dans ces explications. Dennett le sait bien, et c'est pourquoi il est soucieux de dire que les patterns intentionnels ont des pouvoirs causaux. Mais il en prend à son aise en se contentant de leur *accorder* ce statut causal.

Le second problème qui reste pendant est celui de savoir comment marier la perspective intentionnelle, qui est une perspective, comme ne cesse de le souligner Dennett, normative, idéalisante, postulant des explications en termes de ce que les agents *devraient* idéalement faire ou penser, avec une perspective proprement causale. Comment les propriétés purement normatives des contenus peuvent-elles s'associer avec des propriétés causales réelles, qui par définition ne sont pas normatives ?

Davidson a une réponse à cette question. Il soutient que les deux perspectives peuvent se marier parce qu'il est définitoire, ou définitionnel des causes mentales qu'elles soient normatives, alors que la causalité n'a pas cours au niveau des lois physiques :

Les lois strictes ne contiennent pas de concepts causaux, tandis que la plupart, sinon tous les concepts mentaux sont irréductiblement causaux. Une action, par exemple, doit être intentionnelle sous une certaine description, mais une action n'est intentionnelle que si elle est causée par des facteurs mentaux tels que des désirs et des croyances. Les croyances et les désirs sont identifiés en partie par les types d'actions qu'ils sont susceptibles de causer dans les conditions appropriées. La plupart des concepts qui figurent dans les explications du sens commun sont causaux en ce sens. L'accident fut causé par le fait que la route était glissante ; quelque chose est glissant parce qu'il est la cause du fait que les objets appropriés glissent dans des circonstances appropriées. Nous expliquons la raison pour laquelle l'aile d'un avion ne se brise pas quand elle se tord en notant qu'elle est faite en partie de matériaux élastiques ; un matériau est élastique s'il y a quelque chose à son sujet qui est la cause, sous des conditions appropriées, du fait qu'il retourne à sa forme initiale après des déformations. De telles explications ne peuvent pas être précisées pour deux raisons : nous ne pouvons pas expliciter en

25. J. Kim, « Mental Causes and Explanatory Exclusion », in E. Villaneuva, dir., *Information, Semantics and Epistemology*, Blackwell, Oxford, 1990.

détail pourquoi les circonstances sont appropriées, et l'appel à la causalité affine ce qu'une explication complète rendrait manifeste. Les descriptions des objets, des états et des événements qui sont requis pour instantier des lois strictes, sans exception, ne contiennent pas de concepts causaux.

Dans le cas des propriétés causales telles que l'élasticité, le caractère glissant, la malléabilité, ou la solubilité nous tendons à penser, à tort ou à raison, que ce qu'ils laissent inexpliqué peut être expliqué (ou a déjà été expliqué) par les avancées de la science. Nous ne changerions pas le sujet si nous devions abandonner le concept d'élasticité en faveur d'une spécification de la micro-structure des matériaux dans l'aile de l'avion qui le conduit à revenir à sa forme initiale quand elle est soumise à certaines forces. Les concepts et les explications mentales ne sont pas du même genre. Ils font appel à la causalité parce qu'ils sont construits, comme le concept de causalité lui-même, pour extraire, à partir de la totalité des circonstances qui conspirent à causer un événement unique juste ceux des facteurs qui satisfont un certain intérêt explicatif particulier. Quand nous voulons expliquer une action, par exemple, nous voulons savoir quelles sont les raisons de l'agent, en sorte que nous pouvons voir par nous-mêmes ce qui dans l'action attirait l'agent. Mais il serait stupide de supposer qu'il existe des lois strictes qui stipulent que chaque fois qu'un agent a certaines raisons il accomplira une certaine action²⁶.

Le problème d'une telle position est qu'elle ne nous donne aucun moyen de comprendre les relations qui existent entre les concepts mentaux et les concepts physiques. On se contente de dire qu'il *doit* y avoir une correspondance, bien qu'elle ne soit pas nomologique, mais on ne nous dit pas comment on pourrait le comprendre. On doit l'admettre comme un fait inexplicable. Mais s'il y a des régularités aux deux niveaux — mental et physique — comment cette double régularité peut-elle exister autrement que comme une coïncidence, à nouveau selon une sorte d'image occasionnaliste ?

Supposons au contraire que nous puissions comprendre cette correspondance entre les patterns intentionnels et les patterns physiques. Dans ce cas comment éviter de dire qu'il y a certaines lois qui unissent le mental et le physique ? Ces problèmes sont familiers ; ils se posent aussi bien pour le monisme anomal que pour le fonctionnalisme et pour le patternalisme de Dennett. Mais ce dernier ne peut pas se justifier seulement en posant l'analogie entre les structures de la psychologie intentionnelle avec des patterns et en posant que ces patterns doivent avoir des répondants objectifs dans les choses. S'ils ont des répondants objectifs dans les choses, cette réponse des choses à nos patterns doit être *intelligible*, et pas seulement une sorte de fait brut.

J'ai essentiellement, dans cet essai, voulu soulever les problèmes que pose la notion de *pattern* pour la position ontologique de Dennett quant aux états intentionnels, c'est-à-dire pour la version « modérée » de réalisme intentionnel qu'il entend défendre dans « Real Patterns ». Mais cette notion refait surface dans bien d'autres de ses théories. Elle sous-tend sans doute aussi la

26. D. Davidson, *op. cit.*

conception de la conscience qu'il a plus récemment défendue dans *Consciousness Explained* ²⁷. Dans ce livre, Dennett adopte à bien des égards ce que l'on peut considérer comme une théorie éliminativiste de la conscience, en niant, en particulier que la conscience soit un lieu central dans l'esprit « où tout se rejoint », un « théâtre cartésien » où défileraient, pour les *bénéfices* d'un observateur dans l'esprit, les représentations de l'expérience, et en niant qu'il existe des *qualia* ou expériences phénoménales. D'un autre côté, Dennett se récrie quand on l'accuse d'être éliminativiste : il prétend bien avoir fourni une théorie de la conscience comme phénomène *réel*, produit de l'évolution et « machine virtuelle » imposée sur des strates évolutives spécifiques. Comment peut-il concilier les deux idées ? Il est clair qu'il le peut s'il soutient au sujet de la conscience une théorie semblable à celle qu'il défend pour les états intentionnels en général : celle-ci est un pattern, distribué, chaotique et pluriel, produit de coalitions de multiples « démons », dont l'existence n'est pas un fait discernable dans des structures cérébrales, mais qui est « réelle » néanmoins. Les objections que l'on peut adresser à cette idée seraient alors identiques à celles que l'on peut adresser à la notion de pattern appliquée aux croyances. Mais elles doivent être aussi plus spécifiques. Même si Dennett nous dit que les diverses variétés de conscience — la conscience phénoménale, qui produit des *qualia*, la conscience réflexive, qui produit un accès à des représentations — sont elles aussi des patterns, il doit nous dire quelles *sortes* de patterns elles produisent. Et à partir du moment où l'on peut caractériser mieux ces types de conscience, les catégoriser, il devient de plus en plus difficile de soutenir que leur réalité se réduit à celle de trames ou de patterns. Je ne peux développer ce point ici, mais s'il est correct, il conduit à chercher à formuler en tout état de cause une conception *plus* réaliste de ces phénomènes que celle que propose Dennett²⁸.

Université de Caen,
CREA et Institut Universitaire de France

27. D. Dennett, *Consciousness Explained*, Little Brown, New York, 1991, trad. fr. P. Engel, *La conscience expliquée*, Paris, O. Jacob, 1993.

28. J'ai développé ces points dans le chapitre 8 et la conclusion de *Introduction à la philosophie de l'esprit*, Paris, La Découverte 1994, dans « La conscience expliquée ? » à paraître in *Revue Internationale de Psychopathologie*, et dans « Dennett sur la conscience », à paraître in *Neuropsychy*. Alors que je terminais cet article, j'ai eu connaissance de l'article de Norton Nelkin 1994, qui traite des points similaires, et rejoint certaines de mes conclusions. Mais je préfère déférer à une autre circonstance sa discussion.