Document generated on 04/18/2024 5:49 p.m.

Philosophiques

L'empire contre-attaque : le retour de la réduction psychophysique

Pierre Poirier

Volume 27, Number 1, Spring 2000

Le matérialisme contemporain

URI: https://id.erudit.org/iderudit/004969ar
DOI: https://doi.org/10.7202/004969ar

See table of contents

Publisher(s)

Société de philosophie du Québec

ISSN

0316-2923 (print)
1492-1391 (digital)

Explore this journal

Cite this article

Poirier, P. (2000). L'empire contre-attaque : le retour de la réduction psychophysique. *Philosophiques*, *27*(1), 39–62. https://doi.org/10.7202/004969ar

Article abstract

In science, reductions can be relative to specific, well-defined domains, domains that carve nature in a non *ad hoc* way. Relativization to domains is a time-honored way to adjust the conceptual apparatus of theories. And in science, schemes can be defined to describe the global behavior of systems in a way that abstracts from unnecessary (and annoying) detail. I show that the classical argument from multiple realizability to non-reducibility vanishes once the same leeway is granted to psychology. By denying psychology the use of these standard theory-building strategies in science, the argument exhibits an antinaturalist attitude towards psychology, an attitude that may be welcome in some antiphysicalist quarters but that should be shunned in cognitive science.

Tous droits réservés © Société de philosophie du Québec, 2000

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

L'empire contre-attaque : le retour de la réduction psychophysique

PIERRE POIRIER
poirier_pierre@ucdavis.edu
Philosophy Department
University of California at Davis

RÉSUMÉ. — En refusant à la psychologie la latitude accordée aux autres sciences, l'argument concluant à l'irréductibilité des propriétés psychologiques à partir de leur réalisation multiple manifeste une attitude antinaturaliste à l'égard de cette science. En science, il est possible de relativiser les réductions à des domaines bien définis, c'est-à-dire des domaines qui découpent la nature d'une manière non *ad hoc*, et de corriger en conséquence l'appareil conceptuel des théories. Et en science, il est possible de construire des niveaux abstraits et idéalisés permettant la description du comportement global des systèmes, niveaux qui font abstraction de complexités inutiles. Si l'on accorde les mêmes privilèges à la psychologie, la réalisation multiple des propriétés psychologiques ne permet pas d'inférer leur irréductibilité.

ABSTRACT. — In science, reductions can be relative to specific, well-defined domains, domains that carve nature in a non *ad hoc* way. Relativization to domains is a time-honored way to adjust the conceptual apparatus of theories. And in science, schemes can be defined to describe the global behavior of systems in a way that abstracts from unnecessary (and annoying) detail. I show that the classical argument from multiple realizability to non-reducibility vanishes once the same leeway is granted to psychology. By denying psychology the use of these standard theory-building strategies in science, the argument exhibits an antinaturalist attitude towards psychology, an attitude that may be welcome in some antiphysicalist quarters but that should be shunned in cognitive science.

La question de la réduction du psychique au physique est au cœur du problème métaphysique de l'esprit. La théorie classique de la réduction théorique, dont on retrouve la version canonique dans *The Structure of Science* de Nagel[1], devait toutefois remplir une fonction purement épistémique, soit celle d'expliciter les conditions sous lesquelles on peut dire d'une réduction donnée qu'elle est *justifiée*. Ce fait ne devrait surprendre personne : la théorie classique fut développée par les partisans d'une philosophie franchement hostile à la métaphysique. L'analyse néopositiviste des propriétés formelles et sémantiques générales du discours montrait que la métaphysique ne pouvait en aucun cas se prononcer sur les composantes du réel et qu'elle devait être rejetée comme discours insensé lorsqu'elle prétendait le faire. Qui pensait le contraire confondait tout simplement les modes formels et matériels du discours. Cependant, avec la publication coup sur coup des articles de Feigl, « The "Mental" and the "Physical" », et de Smart, « Sensations and Brain

1. Nagel, E., *The Structure of Science*, New York, Harcourt, Brace & World, 1961.

Processes », puis des réponses qu'ils ont suscitées, pensons notamment aux articles de Putnam, « Psychological Predicates », et de Davidson, « Mental Events[2] », la philosophie se redonnait le droit de traiter de questions métaphysiques et donnait par le fait même naissance au problème de l'esprit (*mind-body problem*) tel que nous le connaissons aujourd'hui[3]. Du même coup, la théorie classique de la réduction acquérait une fonction métaphysique : déterminer les constituants ultimes du monde en *faisant connaître*, en anticipation du travail scientifique, quelles théories possèdent les propriétés formelles et sémantiques nécessaires pour que leur réduction à des théories dont nous acceptons déjà les engagements ontologiques soit possible.

Avec la chute du mouvement néopositiviste, l'intérêt des philosophes s'est ainsi déplacé de la justification de la réduction à la détermination de la réductibilité. C'est dans cet esprit que des philosophes favorables à l'idée d'une psychologie intentionnelle qui soit scientifique ont opté pour l'argument antiréductionniste fondé sur la thèse de la réalisation multiple[4]. Il est important de noter que, contrairement à l'usage originel de la théorie classique de la réduction, où le travail philosophique (de justification) est postérieur au travail scientifique à proprement parler, ces nouveaux usages « par anticipation » et métaphysiques de la théorie sont à la merci des développements scientifiques. Ils ne valent que dans la mesure où la science du jour permet une projection *fiable* vers la science de demain. Dans cet article, je montrerai que des développements récents en neurosciences mathématiques (*computational neuroscience*) nous permettent d'ores et déjà d'affirmer que l'argument antiréductionniste fondé sur la réalisation multiple des états psychiques n'est pas valide. Si mon argument est valide, il faudra conclure que la psychologie intentionnelle peut se réduire à la neurologie. Dire qu'elle *peut* se réduire ne signifie pas, évidemment, qu'elle se réduira *de fait*, ou qu'elle se réduira *en entier* ou encore qu'elle se réduira *telle quelle*. Cela signifie, cependant, qu'une position antiréductionniste *a priori* est intenable.

Quelques philosophes ont récemment proposé des nouveaux modèles de la réduction[5] qui, pour des raisons propres à chacun qu'il ne convient pas de

---

2. Feigl, H., « The "Mental" and the "Physical" », *Minnesota Studies in the Philosophy of Science*, 2, 1958, p. 370-497 ; Smart, J. J. C., « Sensations and Brain Processes », *Philosophical Review*, 68, 1959, p. 141-156 ; Putnam, H., « Psychological Predicates », 1967, réédité sous le titre « The Nature of Mental States », dans Putnam, H., *Mind, Language, and Reality*, Cambridge, Cambridge University Press, 1975 ; Davidson, D., « Mental Events », 1970, réédité dans Davidson, D., *Essays on Action and Events*, Oxford, Oxford University Press, 1980.

3. Kim, J., *Mind in a Physical World*, Cambridge (Mass.), MIT Press, 1998.

4. Putnam, « Psychological Predicates » ; Fodor, J. A., « Special Sciences », *Synthese*, 28, 1974, p. 97-115.

5. Bechtel, W. et McCauley, R. N., « Heuristic Identity Theory (or Back to the Future) : The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience », communication présentée au 25e congrès annuel de la *Society for Philosophy and Psychology*, Palo Alto, 1999 ; Bickle, J., *Psychoneural Reduction*, Cambridge (Mass.), MIT Press, 1998 ; Kim, *Mind in a Physical World*.

reprendre ici, brisent l'inférence allant de la réalisation multiple des états psychiques à leur irréductibilité. J'affaiblirai davantage la position antiréductionniste en montrant que, contrairement à ce que l'on prétend depuis trente ans, elle ne découle pas non plus de la conception *classique* de la réduction et qu'il n'est donc pas nécessaire de construire de nouveaux modèles de la réduction pour contrer l'argument de la réalisation multiple. Je ne veux pas laisser entendre par là que ces nouveaux modèles sont construits uniquement pour résoudre ce problème et qu'ils seraient par conséquent inutiles à la lumière des arguments du présent article. Ces modèles traitent aussi d'autres problèmes qui ne trouvent peut-être pas de solution dans la théorie classique de la réduction, pensons par exemple au problème de la causalité mentale.

1. Le fonctionnalisme et l'argument antiréductionniste fondé sur la réalisation multiple

Nous allons reprendre en détail l'argument de la réalisation multiple des états psychiques, car il sera important de préciser la prémisse que nous voulons rejeter. Il est parfois professé et souvent sous-entendu par certains philosophes fonctionnalistes que le fonctionnalisme entraîne l'irréductibilité du psychique, et donc que l'adoption d'une conception fonctionnaliste implique le rejet du réductionnisme[6]. Mais comme le notent Bickle et Kim, l'avantage du fonctionnalisme est qu'il peut s'accommoder d'une variété de positions ontologiques au sujet de l'esprit (y compris du dualisme des substances de Descartes) [7]. Le fonctionnalisme permet l'antiréductionnisme mais ne le recommande ni ne l'impose. Il permet la définition d'une conception cohérente du domaine psychique qui ne soit pas réductionniste, mais cela ne signifie pas que sa vérité dépende de la fausseté du réductionnisme ou qu'elle l'implique[8]. Puisque le fonctionnalisme peut s'accommoder aussi bien du réductionnisme que de l'antiréductionnisme, il faut donc le dissocier de l'une ou l'autre de ces positions. Dans la présente section, nous rappelons comment le fonctionnalisme permet la formulation d'une conception de l'esprit qui soit compatible avec la thèse de la réalisation multiple des états psychiques. Nous soulignons ensuite que bien qu'il soit raisonnable de croire que les états psychiques sont bel et bien réalisés de diverses façons, cela n'est pas de nature à empêcher la formulation des lois de correspondance nécessaires à la justification d'une réduction des théories psychologiques, contrairement à ce que l'on professe.

Un matérialiste peut maintenir que la psychologie offre une description abstraite des états et processus cérébraux servant d'intermédiaires entre la

6. Voir notamment Putnam, « Psychological Predicates », et Fodor, « Special Sciences ».

7. Bickle, *Psychoneural Reductions* ; Kim, *Mind in a Physical World*.

8. L'étape fondamentale de la théorie de la réduction de Kim, *ibid.*, demande par exemple la « fonctionnalisation » des propriétés à réduire.

sensation et le comportement. Mais comment cette psychologie individuera-t-elle ces *états abstraits* du cerveau? Une réponse précise à cette question a dû attendre certaines avancées en sémantique des prédicats théoriques, dont nous nous contentons ici de présenter les grandes lignes : pour toute théorie scientifique *T*, (1) la valeur sémantique de *T* dépend de son degré de confirmation ; (2) la valeur de vérité de chaque énoncé de *T* dépend de celle de tous les autres ; (3) la signification des termes de *T* dépend de leur rôle dans *T* ; chaque terme est ainsi *implicitement défini* par la théorie. David Lewis a formalisé cette nouvelle conception de l'interprétation des termes théoriques[9]. Si *T* est une théorie scientifique confirmée, il est alors possible de l'exprimer au moyen d'un énoncé de la forme « $\exists x_1, ..., \exists x_n\, T\,(x_1,...,x_n)$ » où $x_1, ..., x_n$ sont des variables tenant la place des prédicats de la théorie (transformés en termes singuliers). Au moyen de ces « énoncés de Ramsey », nous pouvons alors dire que le prédicat P_1 de *T*, qui correspond à la variable x_1 de *T* « ramseyifiée », exprime la propriété d'être une chose du type de celles qui jouent le rôle représenté par x_1 dans *T* ramseyifiée. La théorie décrit ainsi la fonction ou le « rôle fonctionnel » joué par les objets constituant l'extension de ses prédicats. P_1 exprime la propriété d'être une chose d'un type ayant un certain rôle fonctionnel, *et tout objet capable de remplir ce rôle fera partie de l'extension du prédicat*. Un rôle fonctionnel est donc une propriété de types d'individus caractérisée par la place qu'ils occupent au sein d'une théorie. Par extension, les propriétés exprimées par de tels prédicats sont appelées « fonctionnelles ». Parce que n'importe quel objet capable de remplir un rôle fonctionnel décrit par une théorie aura par le fait même la propriété fonctionnelle, il s'ensuit que tout prédicat fonctionnel peut, en principe, être réalisé de diverses façons. Cette définition de la nature des prédicats fonctionnels ouvre ainsi la possibilité de leur réalisation multiple.

Cette conception de l'interprétation des termes théoriques s'applique naturellement à la psychologie. Tout comme le prédicat P_1 est implicitement défini par les lois constituant *T*, les prédicats psychologiques comme « croit (que *p*) » seraient implicitement définis par le réseau de lois ou de principes pratiques constituant l'appareil nomologique de la psychologie du sens commun. Les lois ou principes pratiques de cette psychologie détermineraient un rôle fonctionnel pour chaque prédicat exprimant une de ses catégories. Un prédicat « mentaliste » exprimerait ainsi une propriété fonctionnelle : la propriété d'être une chose du type de choses jouant tel rôle fonctionnel déterminé par la psychologie du sens commun. Et, tout comme précédemment, il s'ensuit que tout état (neurologique, électronique, etc.) capable d'occuper le rôle approprié *est* un état psychique. La conception fonctionnaliste permet ainsi d'assigner une signification aux prédicats « mentalistes » malgré l'existence

9. Voir Lewis, D., « How to Define Theoretical Terms », *The Journal of Philosophy*, 67, 1970, p. 427-446 ; Lewis, D., « Psychophysical and Theoretical Identifications », *Australasian Journal of Philosophy*, 50, 1972, p. 249-258.

de multiples référents ne possédant rien d'autre en commun que leur capacité de remplir un certain rôle fonctionnel spécifié par la théorie.

Permettre formellement une chose, ce n'est toutefois pas militer en faveur de son existence. Pour tirer un argument antiréductionniste du fonctionnalisme, il faut montrer que les propriétés psychiques sont bel et bien réalisées de diverses façons et que cette réalisation multiple est de nature à empêcher leur réduction. Commençons par la première démonstration, que nous pouvons diviser en deux volets.

Il est raisonnable de croire qu'un humain et un oiseau peuvent partager un état mental, par exemple le désir de manger un certain fruit visuellement aperçu par l'un et l'autre. Mais il est fort peu probable qu'au même moment l'humain et l'oiseau partagent aussi un état neurologique. L'état neurologique occupant le rôle fonctionnel « désirer manger un fruit » dans l'économie psychique de l'oiseau n'est pas similaire à celui qui occupe le même rôle dans l'économie mentale humaine. Bref, l'état psychique « désirer manger un fruit » se trouve réalisé de deux façons distinctes. Puisqu'il n'y a aucune raison de croire que l'état psychique en question est spécial par rapport aux autres états psychiques, et puisqu'il n'y a aucune raison de croire que l'oiseau est un animal spécial sur le plan psychologique par rapport aux autres animaux, on peut généraliser et affirmer que *les états psychiques sont réalisés de manières distinctes chez les animaux de différentes espèces*. Nous appellerons « interspécifique » cette forme de réalisation multiple entre diverses espèces animales. Il existe aussi une forme plus forte de réalisation multiple que nous appellerons « intraspécifique », car elle prévaut au sein d'une même espèce.

Le cerveau des personnes dont l'hydrocéphalie fut corrigée à la naissance est grandement atrophié et, pourtant, ces personnes sont d'intelligence normale et ne manifestent aucune pathologie cognitive particulière. Il est cependant raisonnable de croire que leurs états psychiques ne sont pas réalisés de la même façon que chez les autres humains. Même chez les humains dont le cerveau est anatomiquement normal, il existe sûrement des différences cérébrales importantes, d'un individu à un autre, au niveau de la structure fine, c'est-à-dire dans le nombre de neurones, le nombre de synapses, la localisation et la force des connexions synaptiques, etc. Puisque tous ces individus peuvent néanmoins partager des états psychiques, il est raisonnable de croire que les états psychiques sont en général réalisés de diverses façons chez les différents individus d'une même espèce. Il paraît aussi raisonnable de croire qu'à ce niveau de description fine, la réalisation multiple intraspécifique ne se limite pas aux différents individus mais qu'elle s'étend jusqu'à l'individu particulier. Nous apprenons vers cinq ans que le père Noël n'existe pas. On dit par ailleurs que chaque individu perd environ un millier de neurones quotidiennement, ce qui signifie que notre croyance s'est maintenue malgré la perte de millions de neurones. Il est donc peu probable que cette croyance soit finement réalisée de la même façon chez un même individu, à l'âge de cinq ans et, par exemple, lorsqu'il a plus de vingt ans. Nous avons de plus appris de nombreuses choses

durant toutes ces années, y compris au sujet du père Noël, et ces connaissances ont certainement modifié la structure fine de nos cerveaux. Bref, il semble raisonnable de croire que les états psychiques d'une personne sont réalisés de diverses façons à divers moments de sa vie[10]. Comme tous ces cas se généralisent, on peut donc affirmer que les états psychiques sont réalisés de diverses façons chez différents types d'organismes, chez différents organismes d'un même type et chez le même organisme à différents moments. La réalisation multiple des états psychiques semble ainsi radicale.

Pour le fonctionnalisme, le seul déterminant de l'identité d'un état psychique réside dans le rôle fonctionnel qu'il occupe. La thèse de la réalisation multiple affirme simplement que des états de types physiques distincts peuvent occuper le rôle fonctionnel définissant un même état psychique. La réalisation multiple des états psychiques est donc tout à fait compatible avec une conception fonctionnaliste de la nature des états psychiques. Ce fait constitue bien sûr un argument empirique important en faveur du fonctionnalisme. Mais certains fonctionnalistes ont cherché à tirer un argument antiréductionniste à partir de la thèse de la réalisation multiple. Autrement dit, selon ce point de vue, cette dernière fournirait non seulement un argument en faveur du fonctionnalisme, mais également un argument contre le réductionnisme. Ainsi, le fonctionnalisme et le réductionnisme entretiendraient une relation d'exclusion mutuelle relativement à la réalisation multiple. Pour justifier cette dernière position, il faut cependant montrer que la réalisation multiple des états psychiques empêche leur réduction à des états neurologiques. C'est ici, et seulement ici, que le réductionniste doit s'inscrire en faux contre le fonctionnaliste antiréductionniste[11].

Complétons d'abord l'argument antiréductionniste fondé sur la réalisation multiple. Remarquons à cet effet que le principe suivant exprime une condition minimale du réductionnisme :

> (1) Toute propriété ne se réduit qu'à une et une seule propriété, ou un et un seul ensemble de propriétés.

Cette condition exprime simplement un aspect du sens du verbe « réduire » tel que l'utilise le réductionniste : si *A* se réduit à *B*, alors il ne peut pas aussi se réduire à *C*. La relation de réduction partage ainsi un trait essentiel avec la relation d'identité. Si *A* est identique à *B* (l'étoile du soir est identique à l'étoile du matin) et si *B* est distinct de *C* (l'étoile du matin est distincte de l'étoile du Nord) alors *A* est aussi distinct de *C* (l'étoile du soir est distincte de l'étoile du Nord). Ce parallèle n'est pas fortuit. La réduction a pour fonction d'élaguer notre ontologie en montrant les identités au sein de l'ensemble des

---

10. Horgan, T., « Nonreductive Materialism and the Explanatory Autonomy of Psychology », dans Wagner, S. et Wagner, R., dir., *Naturalism : A Critical Appraisal*, Notre Dame, Notre Dame University Press, 1993.

11. Churchland, P. S., *Neurophilosophy*, Cambridge (Mass.), MIT Press, 1986 ; Kim, *Mind in a Physical World*.

propriétés et, pour ce faire, elle doit partager la structure logique de l'identité ; bref, dans l'usage réductionniste contemporain, « réduire » doit impliquer « est identique à » et ne peut le faire qu'à condition de partager cet aspect de la structure logique de l'identité[12]. Comment peut-on alors soutenir que le désir de manger un fruit se réalise de deux manières distinctes et que le même désir ne se réduise qu'à un et un seul ensemble de propriétés physiques? La réalisation multiple semble contredire le réductionnisme, dans la mesure où elle entraînerait un conflit entre la condition minimale (1) et des faits comme ceux que nous venons de noter, à savoir :

(2) La réalisation physique du désir que p est *X*.

(3) La réalisation physique du désir que p est *Y*.

Ainsi, l'incompatibilité entre ces trois énoncés montrerait que la réalisation multiple des états psychiques est incompatible avec leur réduction à des états physiques. C'est du moins ce que prétend l'antiréductionniste. Quelle raison a-t-on de le croire?

2. Pourquoi la conclusion antiréductionniste ne découle pas de l'argument fonctionnaliste

Considérons d'abord la réduction interspécifique ; nous discuterons ensuite la réduction intraspécifique. Dans un compte-rendu de l'ouvrage où Putnam a présenté son argument antiréductionniste, Lewis souligne que le conflit entre la réalisation multiple et le principe minimal du réductionnisme, exemplifié ici par l'incompatibilité entre les énoncés (1)-(3), est de même nature que le « conflit » qui résulterait de l'affirmation que les numéros gagnants à la loterie sont 03 et 61[13]. Puisqu'il n'y a qu'un et un seul numéro gagnant à la loterie, comment 03 et 61 peuvent-ils être tous les deux des numéros gagnants? Lorsque le conflit est posé ainsi, il devient évident qu'il disparaîtra aussitôt qu'on précisera le caractère relatif du prédicat « est un numéro gagnant ». Les loteries tirent de nouveaux numéros gagnants sur une base

12. Nagel, dans *The Structure of Science*, souligne que la théorie classique de la réduction n'implique pas nécessairement l'identité mais, comme le remarque Fodor, dans « Special Sciences », la plupart des philosophes réductionnistes ont opté pour une interprétation *forte* de la réduction, qui implique des identités, car elle seule garantit la généralité de la physique. Une réduction qui n'implique pas des identités est compatible avec un dualisme de propriétés ou d'événements qui sied mal avec l'usage contemporain que les réductionnistes font de la théorie de la réduction. Cette différence d'opinion entre Nagel et Fodor s'explique par les intérêts différents qui motivent ces philosophes. Comme nous l'avons noté au début de cet article, l'objectif de ceux qui ont formulé le modèle classique de la réduction n'était pas d'élaguer l'ontologie, puisque les philosophes n'avaient pas, selon eux, à s'occuper de telles questions métaphysiques, mais d'intégrer les théories de niveau supérieur, comme la psychologie, dans la structure de la science unifiée.

13. Voir Lewis, D., « Review of *Art, Mind, and Religion* », *The Journal of Philosophy*, 66, 1969, p. 23-35.

hebdomadaire, de sorte qu'il y a un nouveau numéro gagnant à chaque semaine : le numéro gagnant de la semaine dernière est (ou était) 03, et celui de cette semaine est 61. De la même manière, soutient Lewis, l'apparente incompatibilité entre la réalisation multiple et la condition minimale du réductionnisme se dissipe dès lors que nous précisons le caractère relatif des prédicats psychologiques :

> (1') Toute propriété *d'un type de système* ne se réduit qu'à une et une seule propriété (ou un et un seul ensemble de propriétés) *de ce type de système*.
>
> (2') La réalisation physique du désir que p *dans un système cognitif de type A* est *X*.
>
> (3') La réalisation physique du désir que p *dans un système cognitif de type B* est *Y*.

Le nouveau principe minimal du réductionnisme relativisé à des types de systèmes cognitifs partage toujours la structure logique de l'identité. Si *A*-dans-*X* est identique à *B*-dans-*X*, rien n'empêche que *A*-dans-*Y* soit distinct de *B*-dans-*X*. Le principe permet ainsi d'élaguer l'ontologie tout en respectant la réalisation multiple des prédicats psychologiques. Mais il faut souligner que la relativisation des prédicats psychologiques ne limitera leur extension d'une manière qui convient à la construction de lois psychologiques et de règles de correspondance que dans la mesure où les types de systèmes cognitifs concernés sont ou font partie d'espèces naturelles (*natural kinds*). Sinon, l'antiréductionniste pourra objecter, avec raison, que la relativisation n'est qu'une astuce *ad hoc* servant à prémunir le réductionnisme contre l'objection de réalisation multiple.

Il est assez naturel de laisser à la biologie le soin d'identifier les types de systèmes cognitifs qui pourront servir à la relativisation de la portée des prédicats psychologiques. La sélection naturelle a « construit » les systèmes cognitifs et les a regroupés en types. On peut dès lors penser qu'elle a produit des différences que voudra saisir la psychologie, ou du moins des différences dont celle-ci acceptera de s'accommoder. L'autonomie absolue de la psychologie par rapport aux sciences limitrophes n'est plus une position valorisée, tant en psychologie qu'en philosophie de l'esprit ou des sciences. Suivant cette piste, plusieurs philosophes ont proposé de limiter l'extension des prédicats psychologiques aux espèces biologiques[14]. Les prédicats psychologiques ne seraient pas transférables d'une espèce à une autre, même si l'on parle habituellement comme s'ils l'étaient. Ainsi, bien qu'il existe un usage populaire permettant l'attribution du prédicat « croire que *p* » aux humains et aux bêtes, il est raisonnable de postuler qu'un discours bien policé sur le plan scientifique interdira cet usage permissif du prédicat. Mais s'il s'avère

14. Voir, par exemple, Mundale, J. et Bechtel, W., « Integrating Neuroscience, Psychology, and Evolutionary Biology through a Teleological Conception of Function », *Minds and Machines*, 6, 1996, p. 481-505.

raisonnable de relativiser ainsi les prédicats d'attitudes propositionnelles — en supposant que nous admettions, avec l'antiréductionniste, qu'ils feront partie d'une psychologie scientifique future —, il semble tout aussi raisonnable de rejeter la relativisation d'autres prédicats psychologiques, notamment ceux initialement considérés par Putnam, soit la faim et la douleur. Le réductionniste ne disposerait donc pas des ressources nécessaires pour saisir plusieurs des prédicats fondamentaux de la psychologie.

Mais ce serait là mésestimer les ressources de la biologie. L'ensemble des espèces animales ne forme pas une classe mais une structure liée par des relations de parenté, c'est-à-dire un arbre dont les nœuds sont des espèces et les arêtes des relations de descendance. Puisque le processus à l'origine de ces liens ne « crée » pas les organismes de toutes pièces mais, en général, modifie légèrement l'espèce-mère, la parenté biologique des espèces reflétera une parenté structurelle et fonctionnelle qui aura une incidence sur les sciences prenant ces espèces ou leurs propriétés comme objet. On peut croire, par conséquent, que les prédicats psychologiques se relativiseront à différentes structures évolutionnistes et qu'il existera un rapport entre le degré de parenté entre deux espèces et leur degré de similarité psychologique[15]. Comme le notent Bechtel et McCauley, les espèces apparentées partagent des structures cérébrales et il est raisonnable de faire l'hypothèse que ces structures ont des fonctionnalités identiques ou apparentées (*Heuristic Identity Theory*)[16]. Certaines propriétés psychologiques pourront ainsi se relativiser à l'espèce, d'autres au genre, d'autres à la famille, d'autres à l'ordre, et ainsi de suite[17]. Pour reprendre l'exemple initial de Putnam, on peut penser que la faim se relativise à un très vaste ensemble d'espèces, peut-être à tout le règne *Metazoa*, la nutrition étant une des caractéristiques qui définit le règne animal. En fait, si l'on considère le mode par lequel la sélection naturelle produit de nouvelles espèces, on peut croire que seul un petit sous-ensemble des prédicats psychologiques se relativisera à l'espèce. Croit-on vraiment que chacune des quatorze espèces de pinsons identifiées par Darwin sur l'archipel de Galapagos manifeste des propriétés psychologiques distinctes? Dans ce cas particulier, il sera sans doute préférable de relativiser les prédicats psychologiques à la famille (*Fringillidae*). L'espèce humaine est peut-être l'exception qui confirme la règle puisque la différence anatomique principale qui la distingue des ses plus proches parents (tant ancestraux que contemporains) résulte d'un processus d'encéphalisation qui a multiplié la taille du néocortex. L'espèce humaine se distinguera ainsi par la quantité de propriétés

15. La proposition inverse ne tient pas : la similarité psychologique n'entraîne pas nécessairement une parenté biologique. On parlerait plutôt dans ce cas d'homologies, comme les ailes des oiseaux, des insectes et des chauves-souris.

16. Voir Bechtel et McCauley, « Heuristic Identity Theory (or Back to the Future) : The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience ».

17. Si on utilise la conception cladistique de la classification des organismes, on pourra relativiser les prédicats au niveau du clade, un clade étant un groupe d'embranchements d'animaux ayant une même organisation et une évolution phylétique commune.

psychologiques qui lui sont uniques et sera, par conséquent, celle possédant le plus grand nombre de propriétés relativisées à l'espèce.

L'antiréductionniste objectera peut-être que le prédicat « est un numéro gagnant » a toujours été relatif à un tirage particulier alors que le prédicat « désirer que *p* » n'est pas relatif à une espèce donnée, tant au sein de notre psychologie populaire que de notre psychologie scientifique actuelle. Une telle objection trahit toutefois un penchant pour une conception inacceptable de la science, ou pour une forme de dualisme, avec laquelle l'antiréductionniste ne voudra sûrement pas être associé. L'objection présuppose que les prédicats utilisés dans les diverses sciences sont immuables : s'ils n'étaient pas relatifs dans une théorie *T*, alors ils ne pourront pas l'être dans toute théorie postérieure à *T* qui entend traiter du même domaine. Ce principe, que nous pourrions appeler « le principe de l'immutabilité absolue des prédicats scientifiques », fait cependant de notre antiréductionniste un homme de paille : ce principe n'a jamais été accepté en sciences puisqu'il entrave considérablement son développement. Où serait la physique actuelle si les philosophes avaient réussi à imposer le principe de l'immuabilité absolue des prédicats scientifiques pour empêcher la réforme des concepts de notre physique naïve[18]? Comme le remarque Thagard[19], la science progresse en grande partie par des « révolutions conceptuelles », c'est-à-dire par l'élagage des concepts superflus, la création de nouveaux concepts (soit *de novo*, soit par la division ou la redéfinition de l'extension d'anciens concepts), et enfin la réorganisation des relations genre-espèce et partie-tout entre les concepts.

Nous savons par exemple qu'en dépit de ce que nous enseigne notre physique naïve, il s'avère que la propriété macroscopique de température, disons T_{macro}, se trouve réalisée par de nombreuses propriétés microscopiques distinctes T_{micro1}, T_{micro2}, ..., $T_{micro\cdot n}$. Pour un humain, rien ne ressemble plus à la chaleur de l'air que celle de l'eau et pourtant il s'avère que la réalisation de la température est relative aux états de la matière. La température d'un gaz est l'énergie cinétique moyenne des molécules du gaz, celle d'un solide est l'énergie cinétique maximale des molécules du solide, et celle d'un plasma ne peut être définie en termes d'énergie moléculaire puisqu'un plasma ne contient pas de molécules, mais seulement des atomes ionisés. Le terme général « température » dénote donc plusieurs propriétés macroscopiques distinctes (température d'un gaz, d'un liquide, etc.) qui partagent une certaine similarité superficielle, correspondant chacune à des propriétés microscopiques distinctes. Cela signifie-t-il que la température est irréductible? Au contraire : la réduction de la température est un cas paradigmatique de réduction justifiée[20]. S'il a été possible de réduire la température et ainsi

18. McCloskey, M., « Intuitive Physics », *Scientific American*, 248, 1983, p. 122-130.

19. Thagard, P., *Conceptual Revolutions*, Princeton, Princeton University Press, 1992.

20. Voir Enç, B., « In Defense of the Identity Theory », *The Journal of Philosophy*, 80, 1983, p. 279-298.

de redéfinir le concept de température, c'est parce que l'extension originelle du concept était bien structurée et qu'il a dès lors été possible de redéfinir le concept d'une manière non *ad hoc* permettant la formulation de lois scientifiques. L'extension de T_{macro} peut être ainsi divisée sans reste en cinq sous-ensembles disjoints, et il est raisonnable de croire que toute nouvelle exemplification pourra se grouper sous l'un ou l'autre de ces sous-ensembles sauf, bien sûr, si l'on découvrait un jour un nouvel état de la matière.

Les états de la matière permettent de relativiser naturellement la propriété macroscopique, c'est-à-dire de la découper d'une manière qui respecte la structure fine de la nature. De la même manière, les espèces animales, ou même extraterrestres et robotiques, permettent de relativiser les propriétés psychologiques d'une manière qui respecte aussi la structure fine de la nature. Ici aussi, l'extension du prédicat originel est bien structurée car elle peut être divisée en sous-ensembles disjoints, et il est raisonnable de croire que toute nouvelle exemplification pourra se grouper sous l'un ou l'autre de ces ensembles — sauf si l'on découvre une nouvelle espèce, famille, etc., le cas des extraterrestres ou des robots n'étant pas différent de celui des états encore inconnus de la matière. Les nouvelles extensions créées par cette partition de l'extension du prédicat originel permettront la construction des nouveaux prédicats macroscopiques relativisés qui se réduiront, un à un, à des propriétés microscopiques déterminées.

Le réductionniste peut donc répondre à l'argument de la réalisation multiple *interspécifique* en deux étapes : il montre d'abord que celle-ci demande une modification légitime des prédicats de la psychologie, soit leur relativisation à des espèces, familles, etc., qui permet toujours la formulation de lois psychologiques relativisées ; il rappelle ensuite que la réforme conceptuelle est une étape essentielle du développement de toute science, y compris la psychologie. Une réponse aussi simple, rapide (elle date de 1969, soit deux ans après la parution de l'argument) et décisive aurait dû faire oublier assez rapidement cet argument. Mais force est d'admettre qu'il est toujours aussi populaire trente ans plus tard. Je soupçonne que sa ténacité s'explique ainsi[21] : la réalisation multiple interspécifique est un outil rhétorique servant à introduire une autre forme de réalisation multiple qui, elle, empêche bel et bien la formulation de lois de correspondance. Elle empêche la construction de lois de correspondance en niant l'existence de lois psychologiques générales. Les lois psychologiques doivent en effet subsumer des individus psychologiques mais, si cette autre forme de réalisation multiple s'avère vraie, chaque loi psychologique ne pourra subsumer qu'un et un seul individu (voire un individu à un et un seul moment de son existence). Une telle restriction de l'extension des lois et des

21. Je ne m'attarderai pas ici aux facteurs idéologiques qui faisaient à l'époque, et encore aujourd'hui, de l'antiréductionnisme une position de bon goût ou, comme nous dirions maintenant, politiquement correcte. S'avouer réductionniste en matière psychologique dans le milieu académique contemporain, c'est un peu comme s'avouer un penchant pour la droite, la musique disco ou la viande rouge.

prédicats psychologiques, bien qu'elle soit possible, redéfinirait la psychologie d'une manière inacceptable. Nous verrons cependant que nous pouvons rejeter cette autre forme de réalisation multiple.

Notre antiréductionniste pourrait enfin être tenté de riposter que certaines caractéristiques de la psychologie font que le principe de l'immutabilité des prédicats s'y applique. Pensons, par exemple, au caractère intime de notre rapport à nos états psychiques ou à l'incapacité de connaître le contenu de l'esprit d'autrui. Par cette réponse, l'antiréductionniste s'engage cependant à un dualisme qui postule d'entrée de jeu que l'appareil méthodologique de la science ne s'applique pas en psychologie. Un mécanisme fondamental du développement de toute science a pour tâche de préciser l'extension des prédicats qu'elle utilise à la lumière de nouvelles données ou de nouvelles théories, et quiconque nie, pour les raisons que nous avons notées ou d'autres, l'action de ce mécanisme en psychologie exclut d'entrée de jeu cette discipline du champ des sciences. Il n'est pas impossible, bien sûr, que la psychologie ne soit pas une science, ou pas une science comme les autres, et que le principe de l'immutabilité des principes s'y applique en particulier. Mais on ne peut adopter cette position qu'après avoir laissé agir le mécanisme de réforme conceptuelle et non avant son action, pour nier son applicabilité. Je ne m'attarderai pas sur cette riposte, bien qu'elle soit populaire. Elle fait reposer l'antiréductionnisme sur un antiphysicalisme, et je présuppose que l'antiréductionniste qui nous intéresse ici insiste pour conserver à la psychologie son caractère scientifique. Bref, nous ne sommes pas concernés ici par celui qui achète son antiréductionnisme au moyen d'un chèque antiphysicaliste. Considérons maintenant la réalisation intraspécifique.

La relativisation des prédicats psychologiques aux espèces ou autres groupes biologiques identifiés par leurs propriétés phylogénétiques bloque l'inférence allant de la réalisation multiple *interspécifique* à l'irréductibilité du psychique. Les espèces, ou autres groupes biologiques, permettent une partition empiriquement motivée de l'extension des prédicats psychologiques et il semble qu'il soit tout à fait justifié, sur le plan scientifique, de restreindre les lois psychologiques aux propriétés ainsi relativisées. Mais s'il est raisonnable de limiter les lois psychologiques aux divers groupes biologiques, il n'est pas raisonnable de les limiter aux individus de l'espèce. Le réductionniste peut répondre à l'antiréductionniste que, bien qu'elle soit logiquement possible, sa conception de la psychologie comme science générale dont *toutes* les lois doivent en principe pouvoir s'appliquer tant aux pieuvres qu'aux humains et aux Martiens n'est pas la psychologie scientifique qui se pratique aujourd'hui (ou qu'elle n'en constitue qu'une infime partie[22]). Et il peut aussi objecter que la relativisation des prédicats psychologiques aux groupes biologiques découpe mieux la nature qu'une division incluant les pieuvres, les humains et les Martiens sous une même catégorie. Mais il ne peut du même coup endosser une

22. Mundale et Bechtel, « Integrating Neuroscience, Psychology, and Evolutionary Biology through a Teleological Conception of Function ».

psychologie individualiste qui, bien qu'elle soit elle aussi logiquement possible, ne correspond pas à la psychologie telle qu'elle se pratique aujourd'hui. Et il ne peut pas non plus demander à la psychologie de demeurer aveugle aux lois interindividuelles qui existent manifestement en psychologie. Bref, le réductionniste ne peut répondre à la réalisation multiple *intraspécifique* comme il l'a fait pour la réalisation *interspécifique*, soit en relativisant les prédicats à des groupes plus petits que l'ensemble des systèmes cognitifs possibles. Dans ce qui suit, je montrerai comment on peut rejeter l'inférence de l'irréductibilité du domaine psychique au domaine physique à partir de la réalisation multiple *intraspécifique*.

Revenons à la réduction de la température. Nous avons déjà vu que la relativisation de la température aux divers états de la matière suggère une analogie intéressante avec la relativisation des prédicats psychologiques aux diverses espèces, ce qui permet de répondre au problème posé par la réduction interspécifique. Or, comme le note Bickle[23], la réduction de la température suggère également une analogie pertinente pour le problème posé par la réalisation multiple intraspécifique. À un niveau de description neurologique suffisamment fin, il y a réalisation multiple de la croyance en l'inexistence du père Noël, puisque celle-ci se maintient malgré la perte de millions de neurones. Mais à un niveau de description suffisamment fin, deux températures identiques dans un même gaz seront presque toujours réalisées de manières distinctes, et cela par nécessité nomologique[24]. Lorsque la description physique est suffisamment fine pour saisir les différences individuelles entre deux volumes d'un même gaz, notamment la vélocité et la masse des molécules qu'ils contiennent, deux températures identiques seront presque nécessairement réalisées de manières distinctes. On a donc là, du moins en surface, un cas analogue à celui de la réalisation multiple intraspécifique des propriétés psychologiques. Voyons pourquoi la réalisation multiple intraspécifique de la température ne perturbe pas la thermodynamique des gaz ou la réduction classique de la température, et si nous pouvons en tirer une leçon pour la psychologie.

Les cubes (a) et (b) ci-dessous représentent des volumes identiques d'oxygène contenant seulement trois molécules d'O_2[25]. Les vecteurs représentent les vélocités respectives des molécules. Au niveau de description représenté par le schéma, les deux gaz présentent des propriétés microscopiques distinctes.

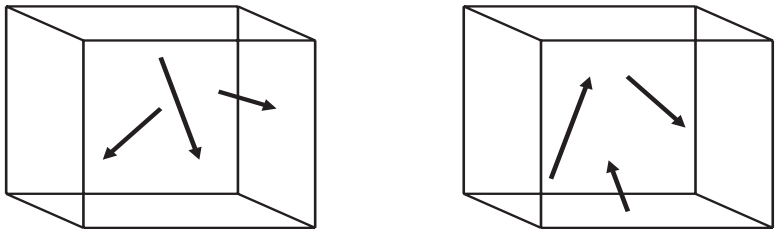
---

23. Bickle, *Psychoneural Reduction*.

24. Les chances que deux températures identiques soient finement réalisées de manière identique sont astronomiquement petites. Cela signifierait que les deux volumes de gaz contiennent le même nombre du même type de molécules ayant toutes la même vélocité (vitesse et direction). Étant donné qu'un mètre cube d'air contient environ 10^{25} particules, il est raisonnable de croire que cette situation ne s'est pas produite depuis le début de l'univers.

25. Ce schéma est de Bickle, *Psychoneural Reduction*, p. 125.

Figure 1



a) b)

La température d'un gaz est proportionnelle à l'énergie cinétique moyenne de ses molécules. La température des deux volumes sera par conséquent identique si l'équation suivante est valide :

$$\frac{\sum_{n}^{1} \tfrac{1}{2} m_a v_a^2}{n} = \frac{\sum_{n}^{1} \tfrac{1}{2} m_b v_b^2}{n} \qquad (1)$$

Puisque nos deux volumes (a) et (b) contiennent un nombre identique de molécules de masse identique (la masse de deux molécules d'un même type est toujours identique : 32 000 *amu* dans le cas d'une molécule de O_2), l'équation se réduit ainsi :

$$v_{a1} + v_{a2} + v_{a3} = v_{b1} + v_{b2} + v_{b3} \qquad (2)$$

Puisque, par hypothèse, seule la composante directionnelle de la vélocité des molécules diffère (les trois vecteurs étant de même longueur), et que cette composante est négligée par la formule (il n'y a pas de facteur la représentant), l'équation se réduit encore :

$$0 = 0 \qquad (3)$$

Nos deux températures sont donc bel et bien identiques mais réalisées de manière distinctes, car les vélocités des molécules diffèrent. Prenons maintenant un seul volume de gaz dont la température demeure constante pour un certain temps. Puisqu'il y aura des collisions et par conséquent des transferts d'énergie cinétique, et puisque les molécules se déplacent (elles ont une énergie cinétique), il s'ensuit que la température identique sera réalisée de manière distincte à n'importe lesquels de deux moments de cette période de temps. Ce cas est plus intéressant que le premier, puisque les collisions amènent des changements de vitesses, c'est-à-dire des changements dans un facteur considéré par l'équation. Mais ces changements de vitesse s'annulent globalement, puisque la vitesse perdue par une molécule sera gagnée par une

autre. Une mesure statistique qui fait la somme des vitesses est donc aveugle à ce genre de transfert énergétique.

Si ces cas de réalisation multiple ne troublent pas les partisans de la réduction thermodynamique de la température, c'est-à-dire l'ensemble de la communauté scientifique contemporaine, c'est que la réduction de la température « ne descend pas[26] » vraiment jusqu'au niveau des propriétés moléculaires des gaz. Elle s'arrête plutôt à un niveau un peu plus élevé où certaines propriétés moléculaires, comme la composante directionnelle de la vélocité, la vibration et la rotation des molécules, etc., ne sont pas considérées. La composante directionnelle est ignorée car, dans un ensemble suffisamment grand, on peut faire l'hypothèse que toutes les directions seront représentées et donc que la direction globale des molécules sera nulle. Mais si les molécules avaient une direction préférentielle, si elles montaient par exemple, il y aurait alors diminution de la pression dans le bas du volume et, donc, diminution concomitante de la température à cet endroit — c'est d'ailleurs ce qui se produit dans l'atmosphère lorsqu'un front froid se déplace. La vibration et la rotation des molécules sont aussi des mouvements moléculaires qui contribuent à la température réelle du gaz. Celles-ci sont toutefois négligées par la formule thermodynamique classique car elles ont un effet négligeable sur la température du gaz. Elles ont cependant leur importance, par exemple, en cryogénie ou lorsqu'on cherche à s'approcher du zéro absolu. La communauté accepte la réduction de la température mais sait pertinemment que celle-ci représente une idéalisation par élimination de facteurs ayant une influence bien réelle mais négligeable. À ce même niveau de description, d'autres propriétés moléculaires ne sont pas négligées mais prises globalement, par exemple la vitesse et la masse, qui sont sommées. C'est pourquoi les changements de vitesse résultant de collisions élastiques (où il n'y a pas perte d'énergie cinétique) s'annulent, une molécule acquérant exactement la vitesse perdue par l'autre. Enfin, à ce même niveau de description, d'autres propriétés encore ne sont considérées que dans une forme idéalisée. Le volume des molécules est postulé être nul (particules ponctuelles), ce qui n'est évidemment pas le cas en réalité. Les collisions, comme nous l'avons déjà remarqué, sont postulées être parfaitement élastiques, etc.

La réduction est établie entre la propriété macroscopique, en l'occurrence la température du gaz, et une propriété introduite à un niveau *abstrait* et *idéalisé* de description mathématique du comportement global des molécules du gaz, soit l'énergie cinétique moléculaire moyenne. Un mètre cubique d'air contient environ 10^{25} particules[27]. Bien que celles-ci se comportent comme un système newtonien et que le concept central de la réduction (l'énergie cinétique) soit un concept mécanique classique (c'est une conséquence directe de la seconde loi de Newton), il est impossible de calculer les interactions au niveau des molécules individuelles. Il est par conséquent

26. Nous poursuivons ici la métaphore de la verticalité inhérente au concept de réduction.
27. Concevoir des molécules comme des particules constitue aussi une idéalisation.

impossible de traiter le gaz comme un système newtonien : il est difficile de calculer les états d'un système newtonien comprenant trois masses comme, par exemple, le soleil, la terre et la lune — ce qui en dit long sur la possibilité de faire le même calcul pour 10^{25} masses. La mécanique statistique et la thermodynamique statistique furent créées pour résoudre ce problème[28].

Bref, la réduction de la température ne se rend pas jusqu'au niveau des propriétés microscopiques des molécules du gaz mais s'arrête au niveau d'une description mathématique plus abstraite et idéalisée. Lorsqu'on descend à un niveau de description plus fin, on observera une réalisation multiple intraspécifique. Ces remarques au sujet de la réduction de la température sont relativement évidentes. Elles montrent qu'on pourra toujours générer une accusation de réalisation multiple intraspécifique si l'on choisit des propriétés suffisamment microscopiques par rapport à la propriété macroscopique dont on entend nier la possibilité de réduction. Elles montrent aussi comment le réductionniste doit répondre à ces accusations : montrer qu'il existe, ou peut exister, un niveau de description *microscopique* où il n'y a pas de réalisation multiple intraspécifique. Car il n'est pas suffisant de remarquer l'illégitimité de l'attaque. Encore faut-il montrer qu'il est possible de définir un niveau de description qui permette de décrire les propriétés des neurones au bon niveau d'abstraction. L'antiréductionniste pourrait en effet répondre : « Merci pour la leçon de physique 101. Mais la neurologie diffère de la physique en ceci : le niveau de description où se pose la réalisation multiple intraspécifique est le seul niveau de description offert par la neurologie, et le seul possible ».

Le débat se réduit donc à la question de savoir s'il existe un équivalent neurologique du niveau intermédiaire de description de la thermodynamique statistique, un niveau de description qui permette de traiter des propriétés des neurones et des circuits neuronaux d'une manière globale, faisant abstraction des détails inutiles ou gênants. Il faut par exemple un niveau de description qui fasse abstraction du nombre de neurones puisqu'une croyance peut se maintenir malgré la perte quotidienne de neurones et puisque deux individus peuvent partager des croyances sans partager le même nombre de neurones. Il faut aussi un niveau de description qui fasse abstraction du détail des interconnexions synaptiques — quel neurone est connecté à quel autre, et avec quelle force — puisqu'une croyance peut se maintenir malgré la formation, la perte et la variation dans la force des interconnexions et puisque deux individus peuvent partager des croyances sans partager une matrice d'interconnexions. Il faut aussi un niveau de description qui fasse abstraction des différences d'enco-

28. Il n'est sans doute pas inutile dans le cadre du débat actuel au sujet de la réduction des propriétés psychologiques de rappeler que ces idées ne furent pas immédiatement acceptées en physique. Plusieurs physiciens voyaient alors d'un mauvais œil et accusaient de « mécaniste » toute tentative de réduire les propriétés électromagnétiques ou thermodynamiques à la mécanique classique (où les corps impliqués seraient des atomes). Je rappelle ces faits pour contrer notre tendance à surestimer les différences entre la physique et les autres sciences. En physique, comme ailleurs, la réduction n'est ni donnée, ni automatique, ni facile.

dage de l'*input* puisque deux individus peuvent partager des croyances au sujet d'un objet malgré la provenance distincte des informations initiales (visuelle et tactile, par exemple). De toute évidence, ce niveau de description devra pouvoir faire abstraction d'un ensemble d'autres différences microscopiques n'ayant pas de répercussion au niveau macroscopique, mais concentrons-nous sur celles que nous venons de nommer.

On pourrait croire que ce niveau de description nous viendra des neurosciences mathématiques (*computational neuroscience*). Les réseaux de neurones artificiels constituent des modèles intéressants, car ils se situent à un niveau intermédiaire entre la psychologie et la neurologie. Ils possèdent à la fois des propriétés microscopiques *semblables* à celles de réels systèmes nerveux et des propriétés macroscopiques *semblables* à celles de réels sujets psychologiques. Par exemple, des systèmes constitués d'unités simples de calcul modélisant certaines des propriétés que l'on retrouve dans les assemblées nerveuses parviennent à départager des énoncés grammaticaux et non grammaticaux[29]. Ces propriétés des réseaux de neurones devraient à tout le moins les rendre intéressants aux yeux de ceux qui s'intéressent à la réduction psychophysique. Comme la thermodynamique statistique, ils semblent offrir un niveau de description intermédiaire ouvrant la possibilité de réduire les propriétés psychologiques aux propriétés d'un système abstrait et partiellement idéalisé qui, sous les conditions appropriées, se réduisent elles-mêmes aux propriétés microscopiques visées. Puisque la réduction est transitive, cette réduction psychophysique en deux étapes constituerait une réduction psychophysique en bonne et due forme, tirant profit d'un pont entre les propriétés macroscopiques de la psychologie et celles, trop microscopiques, de la neurologie.

Mais un réseau de neurones est un système relativement simple, entièrement décrit au niveau microscopique par sa fonction d'activation, son biais, son schème d'interconnexion et sa matrice de coefficients synaptiques[30]. Deux réseaux de neurones possédant la même fonction d'activation, le même biais, le même schème d'interconnexion et la même matrice de coefficients synaptiques sont identiques autant sur le plan microscopique que macroscopique. Cependant, deux réseaux initialement identiques au niveau microscopique peuvent être entraînés à réaliser la même tâche avec un taux de succès identique, c'est-à-dire posséder exactement la même propriété macroscopique et, pourtant, selon le corpus d'entraînement utilisé, posséder des matrices de

---

29. Voir Elman, J. L., « Learning and Development in Neural Networks : The Importance of Starting Small », *Cognition*, 48, 1993, p. 71-99 ; Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., et Plunkett, K., *Rethinking Innateness*, Cambridge (Mass.), MIT Press, 1996.

30. Je ne veux pas présenter ici une introduction générale aux réseaux de neurones artificiels. Le lecteur qui ne serait pas familier avec ces formalismes pourra lire l'excellente introduction de W. Bechtel et A. A. Abrahamsen, *Le connexionnisme et l'esprit : introduction au traitement parallèle par réseaux*, traduction de J. Proust, Paris, La Découverte, 1993.

coefficients synaptiques distinctes. Deux réseaux de neurones artificiels identiques au niveau macroscopique peuvent donc différer au niveau microscopique. Les propriétés macroscopiques des réseaux de neurones se réalisent de diverses façons. Mais, bien qu'il soit abstrait et idéalisé par rapport au niveau de description propre à la neurologie, le niveau microscopique de description propre aux réseaux de neurones artificiels ne peut remplir le rôle d'intermédiaire puisque ces systèmes reproduisent la relation macro/micro que l'on retrouve entre la psychologie et la neurologie. Ce n'est donc pas à ce niveau qu'on trouvera avantage à utiliser ces modèles artificiels.

Les neurosciences mathématiques offrent toutefois la possibilité de définir mathématiquement les propriétés macroscopiques (« psychologiques ») et microscopiques (« neurologiques ») de réseaux de neurones artificiels. Il est par conséquent possible de leur appliquer l'ensemble des techniques mathématiques d'analyse (analyse des tendances centrales, de la variabilité, factorielle, combinatoire, calcul des probabilités, etc.). Ces techniques permettent de suivre à la trace comment les propriétés microscopiques s'organisent pour permettre au système de manifester ses propriétés macroscopiques ; d'établir des métriques variées entre les différents états microscopiques du système et entre ses différents états macroscopiques ; de déterminer si le réseau possède des propriétés plus abstraites, par exemple si tel neurone a tendance à s'activer avec tel autre, etc. Les analyses mathématiques et les possibilités qu'elles ouvrent ne sont pas absentes en psychologie et en neurosciences mais elles sont beaucoup plus difficiles à définir. En neurosciences mathématiques, les propriétés macroscopiques et microscopiques des systèmes sont avant tout des propriétés mathématiques — les systèmes eux-mêmes sont des formalismes — auxquelles on assigne une interprétation fonctionnelle psychologique (dans le cas des propriétés macroscopiques), ou biologique (dans le cas des propriétés microscopiques)[31].

Posons qu'un réseau a été entraîné à discriminer deux types d'objets, c'est-à-dire à répondre de manière différenciée aux objets des deux types, par exemple des échos sonars de mines et de rochers[32]. On pourra regrouper l'activation des divers neurones de la couche cachée du réseau entraîné en un vecteur et représenter celui-ci comme un point dans un espace des configurations[33]. Si l'on reporte l'ensemble des vecteurs d'activation produits par les

---

31. Je ne chercherai pas à évaluer ici la pertinence des interprétations fonctionnelles psychologiques et biologiques ; voir Poirier, P., « Du Stimulus à la Science, neuro-computationnellement », *Actes du XXVII^e Congrès de l'A.S.P.L.F.*, 1999.

32. Voir Gorman, R. P. et Sejnowski, T. J., « Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets », *Neural Networks*, 1, 1988, p. 75-89.

33. Cet espace comprendra une dimension par neurone caché et l'activation du neurone sera représentée par un point dans la dimension appropriée. Ces espaces de configurations sont appelés « espace d'activation » (*activation space*). Ils sont à distinguer des « espaces d'erreur » (*error space*) qui servent à corréler l'erreur en output avec l'état synaptique du système à un moment de son entraînement.

différents objets comme autant de points dans un espace d'activation, on remarquera que les objets initiant une réponse particulière tomberont tous dans une certaine région de l'espace et que plus la réponse est confiante — près de 1, si le spectre de réponse des unités de sortie varie de 0 à 1 — plus le vecteur d'activation tombera près d'un point précis de l'espace. L'entraînement aura amené le réseau à créer une partition de son espace d'activation et à placer un « attracteur » au centre de chacune des deux régions. Plus un input causera une activation des neurones rapprochée de l'attracteur, plus la réponse sera claire et confiante. À l'inverse, plus l'input causera une activation des neurones rapprochée de la frontière de la partition, plus la réponse sera ambiguë et indécise.

Ces différentes propriétés, « posséder une partition », « posséder un attracteur », « tomber dans tel ou tel sous-volume de l'espace », sont des propriétés globales du système. Ce sont aussi des propriétés fonctionnelles générales des réseaux entraînés. Un réseau parviendra à accomplir une tâche s'il réussit à ajuster les coefficients synaptiques de sa matrice de manière à établir une partition appropriée de l'espace d'activation. Notre exemple simple ne demandait qu'une bipartition de l'espace, mais une tâche plus complexe demandera la création d'une partition d'une complexité appropriée. Et il s'avère que les limites « cognitives » d'un système correspondent exactement aux partitions qu'il est incapable de construire, soit parce qu'il ne possède pas suffisamment de neurones cachés, et donc que son espace d'activation est trop petit, soit parce qu'une telle partition n'existe pas, soit parce que le corpus d'entraînement ne contient pas les données appropriées, etc. L'ensemble des propriétés macroscopiques qu'un réseau donné peut manifester est donc déterminé par l'ensemble des partitions de l'espace d'activation qu'il peut créer. La création d'une partition de l'espace d'activation est une propriété générale des réseaux de neurones entraînés pour manifester des propriétés macroscopiques. Or, il est apparu récemment que l'espace d'activation de réseaux distincts sur le plan microscopique mais identiques sur plan macroscopique partagent un ensemble de propriétés géométriques. Ce niveau de description, emprunté à la dynamique, et qui parle d'espace de configuration, de partition de l'espace, de sous-volumes, d'attracteurs, de distance entre attracteurs, de parcours dans l'espace, etc., permet la description de propriétés neurologiques globales à un niveau qui fait abstraction des détails de l'implémentation ou de l'histoire des systèmes particuliers.

Dans une série d'expériences, Aarre Laaksso et Garrison Cottrell[34] ont entraîné dix réseaux de neurones à discriminer des couleurs. Il s'agit là d'un apprentissage très simple, qui n'est qu'une variation de la tâche de reconnaissance d'échos sonars que nous avons notée ci-dessus. L'objet de l'expérience

34. Laaksso, A. et Cottrell, G., « How Can I Know What You Think? : Assessing Representational Similarity in Neural Systems », dans *Proceedings of the Twentieth Annual Cognitive Science Conference*, Mahwah, Lawrence Erlbaum, 1998.

n'était pas de déterminer si les réseaux de neurones peuvent faire ce type d'apprentissage, puisque nous savons depuis longtemps que ceux-ci sont particulièrement habiles à ce genre de tâches discriminatoires. Le réseau devait simplement apprendre à associer des représentations binaires de noms de couleurs (rouge = 10000, jaune = 01000, etc.) à des données provenant d'un spectrophotomètre analysant un échantillonnage standardisé de couleurs[35]. Par exemple, le réseau devait associer le nom « rouge » (c'est-à-dire, 10000) à une donnée provenant d'un spectrophotomètre analysant un échantillon de rouge. Nous avons vu que, d'une manière générale, un réseau réussira cette tâche si sa procédure d'apprentissage parvient à diviser l'espace d'activation en cinq sous-volumes tels qu'un input de rouge amène le réseau à produire le nom « rouge », et ainsi de suite. Les dix réseaux entraînés ne variaient que sous un aspect : le nombre de neurones constituant leur couche cachée. Le premier réseau possédait un neurone caché, le second deux, le troisième trois, et ainsi de suite. Le but de l'expérience était de déterminer si les espaces d'activation de ces dix réseaux entraînés à discriminer des couleurs partageaient des propriétés géométriques globales. L'expérience analysait donc la variation des propriétés géométriques de l'espace (variable dépendante) en fonction de la variation du nombre de neurones cachés disponibles pour accomplir la tâche de discrimination (variable indépendante). Avant de passer aux résultats, précisons la pertinence de cette expérience pour la question qui nous occupe. Soulignons d'abord que les dix réseaux possèdent des propriétés microscopiques distinctes : bien qu'ils possèdent les mêmes couches d'entrée et de sortie, leur couche cachée ne partage pas le même nombre de neurones, et il s'ensuit que le nombre de connexions où distribuer les « connaissances » acquises lors de l'apprentissage varie — par exemple, le réseau à 10 neurones cachés possède 119 connexions de plus que celui à trois neurones cachés, chacune de ces connexions ayant un coefficient synaptique spécifique. La variation du nombre d'interconnexions et de coefficients synaptiques est donc de deux ordres de magnitudes. Nous cherchons ici à déterminer si le réductionniste peut comprendre la réalisation multiple intraspécifique des propriétés psychologiques comme on le fait pour la réalisation multiple intraspécifique des gaz, c'est-à-dire en identifiant un niveau de propriétés qui, bien qu'il soit toujours microscopique, soit suffisamment abstrait et global pour que les différences de réalisations perdent leur pertinence théorique. En permettant la description du comportement global d'ensemble de particules, la mécanique statistique réussit ce tour de force dans le cas de la température. Nous voulons savoir si le vocabulaire de la dynamique des systèmes permet une description similaire des propriétés globales d'un réseau de neurones qui fasse abstraction de certains détails d'implémentation comme le nombre de neurones et le détail précis de leur

35. Anonyme, *Kuopio Color Database*, Lappeeranta University of Technology, 1995 ; http://www.lut.fi/ltkk/tite/research/color/lutcs_database.html.

interconnexion[36]. L'expérience de Laaksso et Cottrell observe justement le comportement de l'espace d'activation en fonction de la variation des propriétés microscopiques du réseau (le nombre de neurones cachés et, *a fortiori*, le détail de l'interconnexion).

Il est apparu au cours de l'expérience que la procédure d'apprentissage ne peut pas diviser de manière appropriée l'espace d'activation des réseaux ne comprenant qu'un ou deux neurones cachés. L'espace d'activation de ces réseaux n'est pas suffisamment « spacieux » pour accommoder les sous-espaces nécessaires à la réalisation de la tâche et, par conséquent, ces réseaux n'ont pas réussi à discriminer les cinq couleurs. Tous les autres réseaux ont su créer une partition appropriée et donc discriminer les couleurs. Ils furent entraînés jusqu'à ce qu'ils réussissent la tâche avec un taux de succès similaire. Pour chacun des huit réseaux capables de discriminer les couleurs, Laaksso et Cottrell ont mesuré la distance euclidienne entre le point causé par un input donné (disons, une représentation d'une donnée spectrophotométrique provenant de l'analyse d'un échantillon de rouge) et les points causés par les autres inputs (jaune, vert, bleu, mauve). Ils ont ensuite fait la somme de ces distances. Cette mesure rend une donnée pour chacun des points de l'espace associé à un input, donnée qu'ils appellent la « distance interpoint » : sa distance globale aux points causés par les autres inputs. On obtient ainsi un ensemble de paires associant des inputs (rouge, etc.) à la distance interpoint entre le vecteur qu'ils causent en espace d'activation et ceux causés par les autres inputs. Si on regroupe toutes les données des huit réseaux, on obtient une matrice où les colonnes représentent les réseaux (donc huit colonnes, une pour chaque réseau capable d'apprendre la tâche), et où les rangées représentent la distance interpoint pour un input donné (une rangée par input). Si on évalue la corrélation entre les distances interpoints pour un input donné, c'est-à-dire la corrélation des nombres au sein d'une rangée, il s'avère que celle-ci est très élevée, soit au-dessus de 0,9. Malgré leurs différences microscopiques individuelles, les réseaux ont appris à répondre à un input donné en produisant un point dans l'espace d'activation dont la distance avec les autres points est identique (ou similaire, $r > 0{,}9$). Nous savions par ailleurs que deux réseaux, initialement identiques, entraînés jusqu'à un taux de succès similaire avec le même corpus, mais où les différentes instances du corpus sont présentés dans un ordre différent, produiront des matrices d'interconnexions différentes et, une fois entraînés, répondront par conséquent en produisant des points différents : les deux réseaux possèdent des attracteurs mais ceux-ci sont situés à différents endroits de l'espace. Cette différence reflète les propriétés microscopiques distinctes des réseaux entraînés. Mais nous savons maintenant que ces points seront équidistants des points produits en réponse aux autres inputs. Grâce à l'expérience

36. De la manière dont elle fut construite, l'expérience de Laaksso et Cottrell fait aussi abstraction de la procédure d'apprentissage, de l'environnement d'apprentissage (l'ordre de présentation des données).

de Laaksso et Cottrell, nous savons aussi désormais que ce résultat est indépendant du nombre de dimensions que contient l'espace d'activation de réseau. La très haute corrélation entre les distances interpoints montre que l'apprentissage a amené les réseaux non seulement à créer une partition appropriée de l'espace d'activation mais à créer un « objet » identique (ou similaire) en espace d'activation, c'est-à-dire un « objet » dont les points saillants sont équidistants, et ce, indépendamment du nombre de dimensions de l'espace dans lequel il se trouve[37]. Il est évidemment difficile d'imaginer visuellement « l'objet » en question. Mais imaginons que nous ayons entraîné un réseau à discriminer seulement deux inputs et que nos dix réseaux aient appris à le faire en associant à chacun un point dans l'espace d'activation. Les résultats de Laaksso et Cottrell suggèrent que les deux points en espace d'activation seront d'une distance égale, ou presque, peu importe le nombre de dimensions de l'espace d'activation du réseau. Il y aura donc une droite de longueur *l* dans un plan, une droite de longueur *l* (approximativement) dans un cube, une droite de longueur *l* (approximativement) dans un hypercube ($n = 4$), une droite de longueur *l* (approximativement) dans un autre hypercube ($n = 5$), et ainsi de suite. Malgré leurs différences microscopiques importantes (variation du nombre de coefficients synaptiques de deux ordres de magnitudes), les réseaux réussissent la tâche en créant un objet similaire, une droite de longueur *l*, en espace d'activation. En comparant seulement la distance entre les points, Laaksso et Cottrell font abstraction de la localisation absolue des points dans l'espace de même que des translations, rotations et réflexions possibles des points dans les divers espaces — différences qui sont causées par l'assignation aléatoire initiale des coefficients synaptiques, l'ordre de présentation des stimuli et le nombre de neurones assignés à la tâche. Par cette même mesure, ils font aussi abstraction de la dimensionnalité des espaces dans lesquels ces points se retrouvent, laquelle représente le nombre de neurones au niveau de la couche cachée.

La réalisation de diverses tâches demande la construction d'objets dans l'espace d'activation et ceux-ci sont créés peu importe le nombre de neurones dont dispose le réseau, si on suppose qu'ils en ont suffisamment. Cela suggère qu'à l'intérieur de certains paramètres, le nombre de neurones et le détail de leur interconnexion ne sont pas des variables neurologiques importantes lorsqu'on s'intéresse à la capacité des réseaux de réaliser une même tâche, tout comme la composante directionnelle de la vélocité des molécules de gaz, ou leur vibration et leur rotation, n'est pas pertinente (à l'intérieur de certains paramètres) à la réalisation de la température. Cela suggère qu'il est possible de concevoir des mesures qui font abstraction de ces variables.

---

37. Rappelons qu'il est toujours possible de placer un objet n-dimensionnel dans un espace contenant un nombre de dimensions égal ou supérieur à n : il est possible de placer un plan ($n = 2$) sur un plan, dans un cube ($n = 3$) ou dans un hypercube ($n > 3$) mais il n'est pas possible de le placer sur une droite ($n = 1$) ou un point ($n = 0$).

Une réponse à l'argument de la réalisation multiple intraspécifique prendra donc la forme suivante. Oui, deux humains peuvent tous les deux croire que le père Noël n'existe pas, même s'ils ne partagent pas le même nombre de neurones et, *a fortiori*, une même matrice d'interconnexion synaptique. Oui, un individu peut continuer à croire que le père Noël n'existe pas malgré la perte quotidienne de neurones et donc la modification quotidienne de sa matrice d'interconnexions. Mais le nombre de neurones et le détail de l'interconnexion ne sont pas des variables neurologiques plus importantes à la réalisation des propriétés psychologiques que la direction des molécules, leur vibration et leur rotation ne le sont à la réalisation de la température. Toute description suffisamment microscopique d'un système sera unique à ce système et impliquera une réalisation distincte d'une propriété macroscopique. Cela montre ou bien qu'il n'existe aucune réduction en sciences, auquel cas l'irréductibilité des propriétés psychologiques n'est qu'une conséquence triviale de l'irréductibilité générale de toutes les propriétés macroscopiques et les scientifiques se leurrent lorsqu'ils croient avoir réduit la température, ou bien que la réduction doit identifier le niveau de description approprié capable de décrire le comportement global des propriétés microscopiques. Tout comme la mécanique statistique permet de décrire le comportement global d'ensembles de molécules d'une manière qui permet la réduction de la température, les neurosciences mathématiques permettent de décrire le comportement global d'ensembles de neurones d'une manière qui permettra la réduction des propriétés psychologiques. Du moins, aucun argument valide ne nous permet désormais de croire le contraire.

3. Conclusion

L'argument inférant l'irréductibilité des propriétés psychologiques à partir de leur réalisation multiple procède d'une attitude foncièrement antinaturaliste à l'égard de la psychologie en lui niant la latitude accordée aux autres sciences. Dans les autres sciences, il est possible de relativiser les réductions à des domaines bien définis, c'est-à-dire des domaines qui découpent la nature d'une manière non *ad hoc*, et de corriger ainsi l'appareil conceptuel des théories utilisées. Dans les autres sciences, il est possible de construire des niveaux abstraits et idéalisés permettant la description du comportement global des systèmes, niveaux qui font abstraction de complexités inutiles. J'ai montré ici que si l'on accorde les mêmes privilèges à la psychologie, la réalisation bel et bien multiple des propriétés psychologiques ne permet pas d'inférer l'irréductibilité du domaine psychique aux propriétés neurologiques ou, plus généralement, physiques des systèmes.

Cela dit, il convient de souligner qu'on commence à peine à comprendre comment construire les niveaux de description appropriés — car il y a en aura certes plus d'un, la psychologie regroupant un amalgame peu unifié de traits humains et animaux — et qu'il revient maintenant aux chercheurs animés du

sentiment réductionniste de construire les niveaux de description appropriés. Les recherches contemporaines, tout comme le contenu du présent article, suggèrent que le niveau de description généralement utilisé pour décrire les autres systèmes complexes, soit la théorie des systèmes dynamiques[38], y tiendra une place importante. Comme le remarque Francis Crick, « lorsque nous en arrivons au point où nous pouvons nous asseoir et pondre la dérivation d'une théorie à partir d'une autre, la partie excitante du véritable travail scientifique est terminée[39] ». Le travail scientifique excitant en psychologie et neurosciences se poursuit aujourd'hui avec la vigueur renouvelée qu'offre la capacité de modéliser mathématiquement les propriétés microscopiques et macroscopiques des systèmes nerveux. Aucun argument ne laisse désormais croire qu'un philosophe ne pourra pas s'asseoir un jour et pondre la dérivation des diverses théories de la psychologie. Malheureusement pour lui ou elle, la partie excitante du travail sera alors terminée.

38. Voir Keslo, J. A. S., *Dynamic Patterns*, Cambridge (Mass.), MIT Press, 1995 ; Tienson, J. et Horgan, T. E., *Connectionism and the Philosophy of Psychology*, Cambridge (Mass.), MIT Press, 1996.

39. Propos rapportés par Patricia Churchland, *Neurophilosophy*, p. 285. Nous traduisons.