Philosophiques

L'émergence, les modèles de réduction et le mental

Jaegwon Kim

Volume 27, Number 1, Spring 2000

Le matérialisme contemporain

URI: https://id.erudit.org/iderudit/004937ar
DOI: https://doi.org/10.7202/004937ar

See table of contents

Publisher(s)

Société de philosophie du Québec

ISSN

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

Explore this journal

Cite this article

Kim, J. (2000). L'émergence, les modèles de réduction et le mental. *Philosophiques*, *27*(1), 11–26. https://doi.org/10.7202/004937ar

Article abstract

A central doctrine of emergentism is the thesis that some properties of wholes are emergent in the sense that they are irreducible to the “basal” properties from which they emerge — that is, they are neither predictable nor explainable in terms of their underlying conditions. To understand and properly evaluate this claim, it is essential that we have at hand an appropriate concept of reduction. This paper first examines the classic Nagel model of inter- theoretic reduction and finds it wanting as a basis for understanding the emergentist claim. Another model of reduction, “functional reduction”, is proposed, and it is argued that this model does provide an appropriate basis for evaluating the claim that certain properties of wholes are emergent. The paper closes with a brief discussion of the question whether *qualia* , or the phenomenal properties of conscious experience, are emergent.

L'émergence, les modèles de réduction et le mental

JAEGWON KIM
Jaegwon_Kim@brown.edu
Department of Philosophy
Brown University

RÉSUMÉ. — Une des doctrines centrales de l'émergentisme est la thèse selon laquelle certaines propriétés d'un tout sont émergentes, en ce sens qu'elles sont irréductibles aux propriétés de base dont elles émergent — c'est-à-dire qu'elles ne peuvent ni être prédites, ni être expliquées à partir de leurs conditions sous-jacentes. Pour comprendre et évaluer cette thèse correctement, il est essentiel que nous disposions d'un concept adéquat de réduction. Nous examinons d'abord le modèle classique de la réduction interthéorique de Nagel, et nous soutenons qu'il ne nous fournit pas une base adéquate pour comprendre la thèse émergentiste. Nous proposons ensuite un autre modèle de réduction, celui de la « réduction fonctionnelle », et nous montrons qu'il constitue une base adéquate permettant d'évaluer la thèse émergentiste. Nous concluons avec une brève discussion de la question de savoir si les propriétés phénoménales d'expériences conscientes, ou *qualia*, sont émergentes.

ABSTRACT. — A central doctrine of emergentism is the thesis that some properties of wholes are emergent in the sense that they are irreducible to the " basal " properties from which they emerge — that is, they are neither predictable nor explainable in terms of their underlying conditions. To understand and properly evaluate this claim, it is essential that we have at hand an appropriate concept of reduction. This paper first examines the classic Nagel model of intertheoretic reduction and finds it wanting as a basis for understanding the emergentist claim. Another model of reduction, " functional reduction ", is proposed, and it is argued that this model does provide an appropriate basis for evaluating the claim that certain properties of wholes are emergent. The paper closes with a brief discussion of the question whether *qualia*, or the phenomenal properties of conscious experience, are emergent.

1. Introduction

L'idée centrale à la notion d'émergence est que, lorsqu'un système composé d'agrégats de matière atteint un certain niveau de complexité organisationnelle, il commence à exhiber de nouvelles propriétés jusqu'alors inconnues, des propriétés « émergentes » — propriétés dont l'occurrence n'aurait pu être *prédite* sur la base des propriétés et relations structurales caractérisant les parties constituantes du système. Cette idée s'accompagne d'une autre idée voulant que l'émergence de telles propriétés ne puisse être *expliquée* à partir des processus sous-jacents (« les conditions de base ») desquels elles émergent. Bref, un tout complexe posséderait de nouvelles propriétés qui sont irréductibles aux propriétés et relations de ses parties. De plus, les propriétés émergentes sont conçues comme étant des propriétés authentiques en

ce sens qu'elles posséderaient des pouvoirs causaux distincts qui leur sont propres. Les systèmes complexes peuvent aussi avoir, et ont effectivement, des propriétés non émergentes, des propriétés réductibles aux propriétés de leurs parties, que les émergentistes britanniques, plus tôt au cours de ce siècle, ont appelées des propriétés « résultantes » ou « conséquentielles ».

On fait habituellement remonter l'idée d'émergence à John Stuart Mill et à sa distinction entre propriétés « homopathiques » et « hétéropathiques », mais la terminologie plus familière de propriétés « émergentes » et « résultantes » est due à G. H. Lewes, un contemporain de Mill[1]. Cependant, la distinction entre ces deux types de propriétés remonte apparemment à beaucoup plus loin, soit aux Grecs et en particulier à Galien[2], ce qui montre bien qu'il y a sûrement quelque chose de naturel et d'intuitif dans l'idée d'émergence. Depuis que ce concept a commencé à faire l'objet d'une attention philosophique particulière, comme ce fut le cas au début du XX^{e} siècle, il a exercé un puissant et persistant attrait sur des penseurs de différents domaines, y compris sur certains scientifiques comme, par exemple, le neurologiste réputé Roger W. Sperry. Je crois que c'est cet attrait intuitif du concept qui a permis de le garder bien en vie au milieu du siècle, malgré les influences marquées du positivisme logique et d'autres courants réductionnistes en philosophie et en science. Il fut un temps, pas si lointain, où l'émergentisme était communément associé à des curiosités métaphysiques largement méprisées, comme le néovitalisme et la doctrine de l'élan vital. Mais le climat intellectuel a radicalement changé au cours des quelques trente dernières années, et le vocabulaire de l'émergence semble être maintenant complètement réhabilité. L'utilisation de l'expression « émergence » et de termes de la même famille se retrouve non seulement dans des textes philosophiques mais aussi dans certains écrits purement scientifiques, surtout dans des domaines comme les sciences cognitives, la psychologie, les neurosciences, la théorie de systèmes, la théorie de la complexité, la théorie de l'auto-organisation, etc. Les termes « émerge » et « émergent » semblent venir assez naturellement aux auteurs dans ces divers domaines, ce qui suggère qu'il y a un contenu substantiel partagé par tous, et cela indépendamment d'un engagement philosophique tendancieux. Aussi, il n'est pas étonnant qu'une perspective largement émergentiste ait fait un retour en force dans les discussions philosophiques portant sur la nature du mental et de sa relation aux processus neurologiques sous-jacents.

Considérons la doctrine du physicalisme non réductionniste qui, depuis les trois dernières décennies, constitue l'orthodoxie relativement au problème corps/esprit (*mind/body problem*). Selon ce point de vue, bien que toutes cho-

---

1. Mill, John Stuart, *A System of Logic*, 8^{e} édition, London, Longmans, Green, Reader and Dyer, 1872, Livre III, ch. iv, paru d'abord en 1843 ; Lewes, G. H., *Problems of Life and Mind*, London, Kegan Paul, Trench, Tubner & Co., 1875.

2. Voir Caston, Victor, « Epiphenomenalisms, Ancient and Modern », *Philosophical Review*, 106, 1997, p. 309-333.

ses capables de mentalité, comme toutes choses en ce monde, soient constituées exclusivement de particules matérielles et d'agrégats de telles particules, les propriétés psychologiques qu'elles exhibent seraient irréductibles à leurs propriétés physiques et elles formeraient ainsi leur propre domaine. De plus, on pense que ces propriétés psychologiques se comportent en accord avec des lois psychologiques irréductibles, qui permettent de produire des explications causales et des prédictions des phénomènes dans le domaine d'investigation scientifique qui leur est propre. Bref, les propriétés psychologiques sont conçues comme des propriétés causales/nomiques d'ordre supérieur qui émergent de configurations complexes de constituants matériels de niveau inférieur, constituants qui eux-mêmes ne posséderaient aucune propriété psychologique. Cette conception du statut de la psychologie et des autres sciences de niveau supérieur est attrayante à plusieurs égards. D'une part, elle renonce au simple dualisme des substances en reconnaissant que les substrats physiques sont nécessaires et suffisants aux propriétés cognitives/psychologiques de niveau supérieur. Et, d'autre part, elle assure à la psychologie une indépendance et une intégrité disciplinaire comme science autonome, sans pour cela devoir accepter l'existence d'entités métaphysiques douteuses, comme les esprits immatériels, les entéléchies ou autres choses semblables. De plus, cette conception peut être généralisée à d'autres sciences spéciales : selon ce point de vue, chaque science spéciale a son propre sujet distinct, constitué par un ensemble de propriétés d'ordre supérieur qui sont irréductibles aux sciences sous-jacentes, et en particulier à la physique élémentaire ; chacune des sciences spéciales cherche des lois causales et tente de formuler des explications fondées sur des lois à l'intérieur de son propre domaine.

La ressemblance entre l'émergentisme et le physicalisme non réductionniste est évidente. Les premiers émergentistes, comme Samuel Alexander, C. Lloyd Morgan et C. D. Broad, auraient trouvé tout cela assez familier et parfaitement naturel, comme un vieil ami perdu et réapparu apparemment inchangé par des épreuves difficiles affrontées au cours des ans.

2. Réduction nagelienne et lois de correspondance

Pour évaluer la situation actuelle impliquant l'émergentisme et le physicalisme non réductionniste, nous devons comprendre clairement le concept de réduction puisqu'il est évidemment au cœur de ces deux doctrines. Comme nous l'avons vu, les propriétés émergentes sont ces propriétés de systèmes complexes qui ne sont pas réductibles aux propriétés et aux relations de leurs constituants — c'est-à-dire que ces propriétés ne sont ni prédictibles ni explicables par des propriétés ou des relations de niveau inférieur. Pour comprendre ce qu'est une propriété émergente nous devons donc comprendre ce qui rend une propriété réductible ou irréductible. Et pour comprendre cela, nous devons comprendre ce qu'est une réduction — c'est-à-dire que nous devons disposer d'un modèle adéquat de la réduction. Comprendre la réduction est

également crucial pour comprendre le physicalisme non réductionniste. Comme nous l'avons vu, la thèse centrale de cette doctrine est essentiellement la même que celle de l'émergentisme : tous les organismes, comme les autres objets de ce monde, sont entièrement faits de fragments de matière, mais certains exhibent certaines propriétés, parmi lesquelles des propriétés mentales, qui ne sont pas réductibles aux propriétés de ces fragments de matière. Mais en fait, comme nous le verrons, ces deux doctrines divergent de façon substantielle, et cela malgré leurs ressemblances superficielles. Plus précisément, nous verrons que ces deux positions sont fondées sur deux notions de réduction passablement différentes et que, par conséquent, du point de vue d'une conception émergentiste de la réduction, certaines versions influentes du physicalisme non réductionniste devraient être considérées comme des théories réductionnistes.

Le consensus antiréductionniste — pour reprendre l'expression bien choisie de Ned Block[3] — qui a régné depuis la fin des années 1960 s'appuie sur une compréhension de la notion de réduction telle qu'elle est entendue dans le modèle classique de la réduction interthéorique développé par Ernest Nagel, dans les années 1950 et 1960[4]. Selon Nagel, réduire une théorie T, comprenant les prédicats primitifs $F_1, ..., F_n$, à une théorie de base T^*, comprenant un ensemble disjoint de prédicats $G_1, ..., G_n$, c'est dériver chaque loi de T à partir des lois de T^* à l'aide de lois de correspondance (*bridge laws*) liant chaque prédicat primitif de T à un prédicat, vraisemblablement complexe, de T^*. Ces lois de correspondance ont la forme suivante :

Pour tout $x_1, ..., x_n$, $F_i(x_1, ..., x_n)$ si, et seulement si, $G(x_1, ..., x_n)$

où G est un prédicat complexe composé de $G_1, ..., G_m$. Ces lois de correspondance, liant les prédicats de la théorie cible à ceux de la théorie de base, sont au cœur de la réduction nagelienne. Cela est évident du fait qu'elles sont non seulement *nécessaires* pour établir des liens déductifs entre les deux théories, formulées dans des vocabulaires hétérogènes, mais elles sont aussi *suffisantes* pour la réduction nagelienne en ce sens que lorsqu'elles sont en place, cela assure une dérivation réductrice des lois-T à partir des lois-T^* *quels que soient les contenus spécifiques de T et de* T^*[5]. Cet aspect de la réduction est important, car il rend possible la discussion des possibilités de réduction avant même que les théories pertinentes n'aient été *complètement*

---

3. Voir Block, Ned, « Anti-Reductionism Slaps Back », *Philosophical Perspectives*, 11, 1997, p. 107-132.

4. Nagel, Ernest, *The Structure of Science*, New York, Harcourt, Brace and World, 1961, ch. 11.

5. La raison pour laquelle les lois de correspondance garantissent la réduction nagelienne est qu'elles permettent la « traduction » des lois de la théorie à réduire par des énoncés formulés exclusivement dans le vocabulaire de la théorie réductrice, c'est-à-dire par des lois de la théorie réductrice. Pour plus de détails, voir Kim, Jaegwon, « Making Sense of Emergence », *Philosophical Studies*, 95, 1999, p. 3-36.

formulées car, bien évidemment, il n'est pas réaliste d'espérer atteindre un tel degré d'achèvement, quelles que soient les théories considérées, dans un domaine assez riche. Bien sûr, nous devons tout de même savoir quelque chose des prédicats des théories en cause. Nous devons d'une part comprendre ces prédicats tels qu'ils sont actuellement utilisés dans les théories, mais nous devons également comprendre comment le vocabulaire de chaque théorie pourrait éventuellement être étendu. Autrement dit, nous avons besoin d'une idée de ce qui fait qu'un prédicat est un prédicat de *T* ou un prédicat de *T**. Par exemple, nous voulons et devons être en mesure de discuter la réductibilité de la psychologie cognitive à la neurobiologie avant de connaître la forme achevée de ces deux théories, car un tel achèvement risque de n'être jamais atteint de notre vivant, et il risque même de ne jamais être atteint tout à fait. Cependant, nous sommes dans une meilleure position vis-à-vis des prédicats. Bien que nos concepts et les prédicats qui les expriment puissent changer selon les vicissitudes de nos théories, ils sont sans aucun doute plus stables que les théories. Après tout, des théories en compétition peuvent être formulées dans le même vocabulaire et, dans la plupart des cas, nous sommes capables de dire quand un prédicat est un prédicat psychologique, un prédicat biologique ou quelque chose d'autre. Et il est possible de construire des arguments réductionnistes ou antiréductionnistes simplement sur la base des caractères généraux des prédicats psychologiques et physiques, indépendamment d'exemples spécifiques.

Il est donc tout à fait compréhensible que le principal argument antiréductionniste, qui fut en grande partie responsable du déclin rapide du réductionnisme au cours des années 1970, ait porté essentiellement sur la disponibilité de lois de correspondance liant les prédicats psychologiques aux prédicats physiques (ou les propriétés psychologiques aux propriétés physiques, pour parler en termes métaphysiques). Il s'agit, bien sûr, du célèbre argument de la réalisation multiple qui, comme nous le savons, a eu une très grande influence[6]. Cet argument ne concernait pas — ou à tout le moins pas directement — la question de savoir si les lois physiques ou biologiques pourraient suffire à produire des lois psychologiques de façon purement déductive ; l'argument était totalement indépendant des contenus particuliers des théories psychologiques et physiques, quels qu'ils aient pu être. Mais il mettait plutôt l'accent sur la question suivante : *Existe-t-il des lois de correspondance psychophysiques?* Et, comme nous le savons, l'argument de la réalisation multiple répond négativement à cette question. Chaque propriété psychologique a un nombre indéfini de réalisateurs divers au niveau physique, et il est impossible de déterminer, à l'aide d'une loi de correspondance ayant la forme notée plus haut, quelle propriété physique serait le corrélat

6. Selon moi, cet argument a également été très mal compris. À ce sujet, voir Kim, J., « Multiple Realization and the Metaphysics of Reduction », dans Kim, J., *Supervenience and Mind*, Cambridge, Cambridge University Press, 1993.

d'une propriété psychologique[7]. Cet argument est habituellement complété par un argument additionnel montrant qu'il serait inadéquat de formuler une loi de correspondance en utilisant, comme corrélat physique, une disjonction de tous les réalisateurs possibles d'une propriété mentale[8].

Ces considérations conduisent assez naturellement à une conception des propriétés psychologiques comme étant des propriétés formelles abstraites qui sont indépendantes de leur composition matérielle ou du mécanisme responsable de leur implémentation. Elles conduisent également à une certaine conception de la nature de la psychologie comme science. Brièvement, l'idée est que les propriétés psychologiques font abstraction des détails physiques/compositionnels concrets de ce qui les réalise, et que l'étude scientifique de ces propriétés, des lois qui les gouvernent et des explications des phénomènes psychologiques peut être menée sans tenir compte des implémentations ou mécanismes physiques qui les sous-tendent. C'est en ce sens que la psychologie et les autres sciences « spéciales » ont été dites autonomes relativement aux sciences physiques/biologiques sous-jacentes.

En résumé, on peut dire que le principal argument contre la réduction psychophysique, qui a engendré les formes de physicalisme non réductionniste présentement populaires, est l'argument de la réalisation multiple, et que celui-ci repose essentiellement sur l'affirmation qu'il n'existe pas de lois de correspondance psychophysiques, ce qui rendrait impossible une dérivation nagelienne de lois psychologiques à partir de lois physiques. Toutefois, ce que l'on note plus rarement ce sont les présuppositions et les implications physicalistes/réductionnistes de cet argument. En effet, dans cet argument, on retrouve implicitement la proposition que les propriétés mentales sont physiquement réalisées, et même la proposition plus forte que chaque fois qu'une propriété mentale est exemplifiée dans un organisme (ou un système), ce serait en vertu du fait que certaines propriétés physiques de l'organisme la réalisent, ou fournissent un mécanisme l'implémentant dans cet organisme. En ce sens, n'importe quelle forme de non-réductionnisme qui s'inspire de l'argument de la réalisation multiple est en fait une doctrine physicaliste de façon non triviale. Et, comme nous le verrons, une telle approche mène ultimement au réductionnisme, une position diamétralement opposée à celle de l'émergentisme. Selon mon point de vue, le fonctionnalisme, tel qu'il est aujourd'hui largement répandu, est un exemple d'une telle position. Cette façon de traiter le problème de la relation corps/esprit a été communément acceptée comme étant une approche non réductionniste[9], mais comme nous

---

7. C'est Jerry Fodor qui a proposé la version la plus influente de cet argument. Voir Fodor, J. A., « Special Sciences », réédité dans Fodor, J. A., *Representations*, Cambridge, Cambridge University Press, 1981. (Cet article a paru initialement en 1974.)

8. Sur cette question, le lecteur intéressé pourra consulter les articles de Kim et de Fodor mentionnés aux notes 6 et 7.

9. Ce ne sont pas tous les fonctionnalistes qui ont endossé cette thèse. David Armstrong et David Lewis sont parmi les rares d'entre eux qui ne l'ont pas fait. L'attitude des fonctionnalistes

le verrons dans ce qui suit, dès lors que nous adoptons un modèle de la réduction plus adéquat que le modèle nagelien, il apparaît clairement que le fonctionnalisme est en fait une forme de réductionnisme.

3. Un *desideratum* de réduction, et la réduction fonctionnelle

La psychologie est-elle réductible, au sens de la réduction nagelienne, à la neurobiologie ou à une autre théorie non psychologique sous-jacente? Avant de tenter de répondre à cette question, nous devons d'abord répondre à une autre question concernant la signification philosophique d'une réduction nagelienne de la psychologie. Si nous supposons que la psychologie a effectivement été réduite en ce sens, qu'est-ce que cela montrerait? La réponse est que cela montrerait bien peu de choses, car selon les théories dualistes classiques concernant la relation corps/esprit — comme par exemple la théorie spinoziste du double aspect, la doctrine de l'harmonie préétablie de Leibniz, le parallélisme ou le monisme neutre — les lois psychophysiques seraient disponibles en abondance. En fait, l'existence de telles lois n'est même pas exclue par le dualisme cartésien, bien qu'elle soit probablement exclue par la théorie spécifique que Descartes a effectivement soutenue — du moins pour ce qui concerne certains états dotés de contenu (soit les « pensées rationnelles »). Cela signifie que la réductibilité nagelienne du mental est non seulement compatible avec ces théories dualistes, mais qu'elle est en fait impliquée par plusieurs d'entre elles. Si la réduction psychophysique doit être comprise sur la base du modèle de la réduction nagelienne, le réductionnisme n'aurait donc pas de conséquences métaphysiques intéressantes. Mais alors, pourquoi devrait-on se préoccuper de la réduction de l'esprit au corps ou de réductionnisme?

Cela jette un doute sur l'adéquation de la réduction nagelienne comme base de discussion du problème corps/esprit, et sape la pertinence philosophique des arguments antiréductionnistes qui reposent sur ce modèle.

Considérons la réduction nagelienne à partir du point de vue émergentiste. La principale question que les émergentistes avaient à l'esprit, à propos de la relation entre les propriétés de niveau supérieur et leurs conditions de base, était la suivante : l'occurrence d'une propriété de niveau supérieur peut-elle être prédite *uniquement* à partir d'informations concernant le niveau de base? La question concernant les phénomènes mentaux serait donc la suivante : l'occurrence de la douleur, par exemple, peut-elle être prédite uniquement à partir de la connaissance de la physiologie du cerveau? Cette première question s'accompagne également d'une question explicative : peut-on expliquer l'occurrence d'une propriété émergente uniquement sur la base

---

envers le réductionnisme a été marquée par une ambivalence profonde et une instabilité incessante, et cela même si ce sont plutôt les vues de certains antiréductionnistes qui ont prédominé. J'espère que le présent article permettra d'expliquer et de résoudre cette ambivalence.

d'une connaissance du niveau de base? Peut-on expliquer pourquoi vous ressentez de la douleur par exemple, plutôt qu'un picotement, uniquement sur la base de faits physiques/biologiques à propos de vous? Si une réduction d'une théorie d'un domaine *M* à une théorie de base ayant pour domaine *P* ne fournit pas de prédictions ou d'explications d'une occurrence des phénomènes de *M* sur la base de faits dans *P*, que peut donc nous procurer une telle réduction?

Ces considérations nous amènent à la suggestion suivante : si la notion de réduction doit avoir un rôle significatif dans les débats sur la relation corps/esprit, alors elle doit satisfaire le *desideratum* suivant :

> Si une propriété d'ordre supérieur *P* est réductible à un niveau inférieur *L*, alors l'occurrence de *P* doit être prédictible et explicable uniquement sur la base d'informations concernant les faits de niveau *L*.

Si nous avions un modèle de la réduction satisfaisant cette exigence, cela pourrait motiver une définition des propriétés émergentes comme étant ces propriétés de niveau supérieur qui ne sont pas réductibles, selon ce modèle, à des faits d'un niveau inférieur quelconque[10]. Une telle compréhension de la notion d'émergence a pour conséquences heureuses qu'une propriété est émergente relativement à ses conditions de base seulement si elle ne leur est pas réductible, et que l'émergentisme est vrai d'un ensemble de propriétés seulement si le réductionnisme est faux de ces propriétés.

Selon les émergentistes, il existe bien sûr certaines propriétés d'un tout qui *sont* effectivement prédictibles sur la base des propriétés de leurs parties. Ils citeraient en exemple la masse : la masse de cette table peut être dérivée, logiquement ou mathématiquement, de la masse de chacune de ses parties, disons le dessus de la table et sa structure[11]. De plus, nous pouvons expliquer pourquoi la table a cette masse à partir d'informations à propos de la masse de ses parties. Mais, toujours selon les émergentistes, diverses propriétés de composés chimiques, comme par exemple la transparence et la viscosité de l'eau, ne seraient pas dérivables, et donc non prédictibles à partir des propriétés de leurs atomes constituants ; c'est-à-dire que les configurations structurales au niveau des molécules d'eau ne seraient pas dérivables à partir des propriétés des atomes d'hydrogène et d'oxygène considérées isolément. Cette affirmation des émergentistes paraît plausible, du moins à première vue, en raison de l'absence des concepts de transparence et de viscosité dans la physique atomique. Et, toujours selon ce point de vue, cela vaudrait également pour la mentalité et la conscience, ce qui veut dire qu'on peut tout savoir des processus physiques et biologiques se déroulant dans un organisme sans pour

---

10. Suivant une pratique répandue, nous utilisons ces notions impliquant différents « niveaux » en un sens intuitif — particulièrement les notions de « supérieur » et « inférieur » telles qu'appliquées aux niveaux. Ces notions sont discutées de façon plus détaillée dans Kim, J., « The Metaphysics of a Layered World », à paraître.

11. Il est douteux cependant que cette thèse émergentiste soit vraie. L'additivité de la masse serait aujourd'hui considérée comme une propriété empirique et contingente.

autant connaître, de ce fait, dans quel état de conscience il se trouve, si état de conscience il y a — c'est-à-dire sans savoir s'il a ou non une expérience consciente et, si c'est le cas, quel en serait le caractère qualitatif.

Un trait intéressant du *desideratum* sur la réduction, que nous avons formulé plus haut, est qu'il n'exige pas que la propriété *M*, visée par la réduction, soit identifiée à une propriété ou même à un groupe ou à une disjonction de propriétés au niveau de base. (En fait, il n'exige même pas qu'il y ait des lois de correspondance nagelienne.) L'explication ou la prédiction d'une exemplification de *M*, en une occasion particulière, peut être « locale » au lieu de viser à être « globale » en ce sens qu'elle peut recevoir une explication réductrice qui en appelle à un mécanisme ou à une structure sous-jacente impliquée dans *M* en cette occasion particulière, alors qu'une autre exemplification de *M* pourrait recevoir une explication invoquant un mécanisme complètement différent. La raison en est que rien ne garantit qu'un seul et unique mécanisme ou une seule et unique structure soient sous-jacents à toutes les exemplifications actuelles et possibles de *M* ; en fait, il est plausible que le contraire soit la norme. Ce qui, bien entendu, n'est qu'une autre facette du phénomène de la réalisation multiple. Cela signifie qu'un modèle de réduction peut satisfaire le *desideratum* sans exiger de réductions type/type, et donc que les modèles qui satisfont ce *desideratum* peuvent le faire tout en conciliant la réalisation multiple des propriétés réduites.

Je vais maintenant esquisser un modèle de réduction qui satisfait le *desideratum* formulé plus haut. C'est ce que j'ai appelé la « réduction fonctionnelle », qui consiste dans les trois étapes suivantes :

> *Étape 1*. Fonctionnaliser la propriété visée par la réduction — c'est-à-dire donner une caractérisation de la propriété par son rôle causal qui est spécifié dans les termes des propriétés au niveau de base.

Une telle caractérisation est appelée une *définition fonctionnelle* de la propriété en question. La première étape dans la réduction du gène est de le comprendre fonctionnellement, c'est-à-dire dans les termes du « rôle causal » qu'il est supposé remplir. Ainsi, on explique le gène comme ce mécanisme qui, dans un organisme, accomplit la tâche d'encoder et de transmettre l'information génétique.

> *Étape 2*. Identifier la ou les propriétés (ou le ou les mécanismes) présents au niveau de base, qui accomplissent les tâches causales spécifiées dans le ou les systèmes examinés — c'est-à-dire trouver le réalisateur de la propriété fonctionnalisée pour les systèmes examinés.

Au contraire de l'étape 1, l'étape 2 est carrément dans le domaine de la recherche empirique/théorique. Les philosophes n'ont aucune contribution à apporter à cette étape. Notons que, étant donné le phénomène de réalisation multiple, l'étape 2 sera normalement exécutée cas par cas, car il arrivera rarement que la recherche scientifique vise et réussisse à découvrir tous les réalisateurs nomologiquement possibles implémentant, à un niveau inférieur, la

propriété ciblée — ou qu'il y ait un unique mécanisme sous-jacent qui puisse être mis au jour. Il est plus plausible que les chercheurs tenteront d'identifier le ou les réalisateurs impliqués dans une espèce particulière ou dans un groupe de structures qui présentent un intérêt pour leur programme de recherche. Ainsi, pour les organismes terrestres, les molécules d'ADN furent identifiées comme porteurs et transmetteurs de l'information génétique. Mais il est nomologiquement possible que différentes molécules fassent ce travail biologique dans différentes espèces d'organismes terrestres. Qui plus est, dans des mondes possibles où prévaudraient des lois physiques différentes des nôtres, des molécules autres que celles d'ADN pourraient accomplir cette tâche.

> *Étape 3*. Développer une théorie qui explique comment les réalisateurs de la propriété cible font pour accomplir la tâche causale spécifiée.

Nous voulons une théorie qui explique comment les molécules d'ADN font pour encoder et transmettre l'information génétique des parents à leur progéniture. Une compréhension réductrice du gène — c'est-à-dire une compréhension du gène en termes de processus sous-jacents — devrait, semble-t-il, inclure une telle théorie. De toute façon, les étapes 2 et 3 seront presque certainement accomplies en collaboration : on peut s'attendre à ce qu'elles constituent deux parties, ou deux aspects, d'un même programme de recherche. Il est peu probable que la recherche puisse identifier les réalisateurs d'une propriété scientifiquement significative sans avoir recours à une théorie explicative des phénomènes sous-jacents mis en jeu. Au-delà de la caractérisation causale donnée par la définition fonctionnelle de la propriété cible, une théorie explicative en développement aidera à circonscrire et à désigner les endroits où les réalisateurs pourraient se trouver[12].

4. La satisfaction du *desideratum* de réduction

Nous allons maintenant voir comment notre *desideratum* sur la réduction — le réquisit émergentiste d'explication et de prédiction — peut être satisfait si

---

12. Ces idées proviennent, à l'origine, des certains écrits de David Armstrong et de David Lewis portant sur une version du matérialisme qui, à l'époque, était connue sous l'appellation de « Central State Materialism », bien qu'ils ne les aient pas proposées explicitement comme fournissant un modèle de la réduction. Voir Armstrong, D., *A Materialist Theory of Mind*, London, Routledge & Kegan Paul, 1968 ; ainsi que Lewis, D., « An Argument for the Identity Theory », *Journal of Philosophy*, 67, 1970, 203-211. Certaines idées similaires ont récemment refait surface, étant cette fois plus clairement identifiées à la réduction et à l'explication réductrice chez Joseph Levine dans son article « On Leaving Out What It's Like », dans Davies, M. et Humphreys, G., dir., *Consciousness*, Oxford, Blackwell, 1993. Ce que je propose ici est que ce modèle, et non pas le modèle de Nagel, constitue un modèle adéquat de la réduction, et cela non seulement pour ce qui concerne le débat sur le problème corps/esprit, mais pour la réduction en général. Pour une discussion plus élaborée, voir Kim, J., *Mind in a Physical World*, Cambridge, MIT Press, 1998 ; et « Making Sense of Emergence », *Philosophical Studies*, 95, 1999, p. 3-36.

une propriété est réductible au sens du modèle de la réduction fonctionnelle que nous venons d'esquisser.

Considérons d'abord la prédiction de l'occurrence d'une propriété *M* fonctionnellement réduite : est-ce qu'un système donné *S* exemplifie *M* au temps *t*? Pour répondre affirmativement à cette question, en accord avec la caractérisation fonctionnelle de *S*, *S* devrait exemplifier, à *t*, une propriété *P* remplissant un certain rôle causal *R*. Comme nous le supposons dans ce cas, une réduction fonctionnelle a été accomplie, et nous supposons donc comme étant acquise la connaissance du réalisateur de la propriété *M* dans des systèmes comme *S*, à savoir *P* qui est une propriété ou un mécanisme du niveau de base. La question prédictive est donc la suivante : est-ce que *S* exemplifie *P* au temps *t*? Manifestement, il s'agit là d'une question qui peut recevoir une réponse au niveau de base, ce qui signifie que les occurrences de *M* peuvent effectivement être prédites à partir d'informations concernant uniquement le niveau de base. (Il peut y avoir plus d'un réalisateur de *M* dans des systèmes comme *S*, mais cela ne fait aucune différence.)

Tournons-nous maintenant vers l'explication de l'occurrence d'une propriété *M*. Pourquoi le système *S* exemplifie-t-il *M* au temps *t*? Encore une fois, pour que *S* exemplifie *M* à *t*, *S* doit exemplifier un réalisateur de *M*, à *t*. *S* exemplifie donc *P* à *t*, et *P* est un réalisateur de *M* — *P* satisfait le rôle causal spécifié pour *M*. Et c'est pourquoi *S* exemplifie *M* à *t*.

Considérons l'exemple suivant : Jean éprouvera-t-il de la douleur au temps *t*? Éprouver de la douleur est pour lui, par définition, être dans un état (c'est-à-dire exemplifier une propriété) susceptible de causer des cris et des hurlements, et un comportement de fuite. Chez les humains, l'activation de fibres *C* est le réalisateur de la douleur. La question est donc de savoir si les fibres *C* de Jean seront activées au temps t. Et ce sont des faits et des lois au niveau physique/biologique qui permettront, en principe, de répondre à cette question.

Pourquoi Jean éprouve-t-il de la douleur au temps *t* (au lieu d'un picotement, par exemple)? Avoir de la douleur c'est se trouver dans un état susceptible de causer des cris et des hurlements, et un comportement de fuite. Au temps *t*, les fibres *C* de Jean sont activées et l'activation des fibres *C* est un réalisateur de la douleur chez les humains. (Et la stimulation des fibres *C* n'est pas le réalisateur de picotements chez les humains.) Voilà pourquoi Jean éprouve de la douleur, et non pas un picotement, au temps *t*.

Il me semble que, dans ces deux cas, nous sommes en présence d'une véritable prédiction et d'une véritable explication. De plus, celles-ci sont *réductrices* en ce sens que les ressources explicatives et prédictives qui sont mises à profit proviennent exclusivement du niveau réducteur de base. C'est une définition qui nous fournit les relations réductrices entre le niveau de base et la propriété cible, en l'occurrence une caractérisation fonctionnelle de *M* en termes de relations causales impliquant des éléments du niveau de base.

Comparons cela à la situation impliquant une réduction nagelienne de *M*. Supposons qu'une théorie à laquelle *M* appartient ait été réduite à une théorie de niveau de base à l'aide de lois de correspondance incluant celle-ci :

M T

où *T* est une propriété de niveau de base. Étant donné une telle réduction, comment prédire ou expliquer, de façon réductrice, une occurrence de *M*?

Notons que la propriété *M* n'a pas d'occurrence dans la théorie de base initiale, et que c'est seulement la théorie de base enrichie — à savoir la théorie initiale plus les lois de correspondance — qui inclut *M*. En fait, c'est par le biais de la loi de correspondance notée plus haut que *M* apparaît dans la théorie enrichie. Ainsi, la seule façon de prédire des occurrences de *M* serait d'utiliser cette loi de correspondance, essentiellement de la manière suivante :

Le système *S* exemplifiera *T* à *t*

M T

Donc, *S* exemplifiera *M* à *t*.

Notons cependant que cette prédiction ne satisfait pas le *desideratum* sur la réduction. Car, bien que ce soit une prédiction inductive parfaitement acceptable, elle s'établit sur une base comprenant la loi de correspondance. Et celle-ci véhicule de l'information préalable au sujet de la corrélation entre les occurrences de *M* et celles de *T*, ce qui constitue une connaissance présupposant des observations préalables de *M*. De telles prédictions d'occurrences de *M* ne sont donc pas fondées *uniquement* sur la base d'informations concernant le phénomène et les lois au niveau de base et, par conséquent, elles ne satisfont pas le *desideratum* sur la réduction.

Les émergentistes savaient bien que nous faisons couramment des prédictions inductives d'états conscients sur la base de corrélations observées entre ces états et des états comportementaux ou neuronaux. Mais la question qui avait pour eux un intérêt fondamental était plutôt la suivante : étant donné tout ce que l'on peut savoir au sujet du comportement et du système neuronal d'un organisme, et étant donné cette *seule* information, est-il possible de prédire ou d'inférer que cet organisme éprouve de la douleur ou qu'il est conscient? Leur réponse était que cela est impossible. Il ne suffirait pas de répliquer que l'on peut inférer que l'organisme doit éprouver de la douleur parce qu'il a un état cérébral similaire à celui des humains lorsque ceux-ci éprouvent de la douleur. Car une telle réponse fait encore une fois appel à une connaissance préalable des corrélations entre la douleur et les états du cerveau.

Nous pouvons donc conclure que les réductions se conformant au modèle de Nagel, contrairement aux réductions fonctionnelles, ne fournissent ni des prédictions réductrices ni des explications réductrices. Elles ne satisfont tout simplement pas le *desideratum* sur les réductions que nous avons formulé plus haut.

5. La conscience est-elle réductible ou émergente?

Revenons maintenant à la réduction de l'esprit. Est-ce que la mentalité — c'est-à-dire l'ensemble des diverses propriétés mentales — peut être réduite? Et si elle le peut, quelle serait sa base de réduction? Le cerveau et le comportement sont bien sûr deux candidats assez évidents à ce titre. Le behaviorisme logique ou analytique fut une tentative pour réduire les propriétés mentales en les définissant en termes de comportements ou de dispositions comportementales. Mais il est largement admis que cette tentative a échoué. Comme on l'a fréquemment souligné, les concepts mentaux ou les propriétés mentales ne peuvent tout simplement pas être définis en termes de dispositions comportementales.

La théorie de l'identité type/type de Smart et Feigl[13], qui a brièvement régné à la fin des années 1950 et au début des années 1960, peut être considérée comme une théorie réductionniste modelée sur la conception nagelienne de la réduction, mais à une différence importante près. Leur suggestion était de remplacer les *bi-conditionnels* psychophysiques constituant les lois de correspondance, qui formulent des corrélations uniformes entre les deux domaines séparés du mental et du physique, par des *identités* psychophysiques. Ainsi, la loi de correspondance corrélationnelle « La douleur se produit chez un organisme au temps *t* si, et seulement si, il a un état cérébral *X* au temps *t* » serait remplacée par une loi d'identité, jouant le rôle du principe de correspondance, à savoir : « La douleur = l'état cérébral *X* », et les autres lois de correspondance seraient remplacées de façon similaire. Mais ces identités, connues sous le nom d'« identités théoriques »[14], étaient considérées comme étant empiriques et contingentes. Or il est clair que toute réduction reposant sur l'intermédiaire de lois de correspondance empiriques, qu'il s'agisse de corrélations ou d'identités, ne parvient pas à satisfaire le *desideratum* sur la réduction. Ce point est particulièrement important dans la mesure où une théorie de l'identité type/type, à la Smart/Feigl, a récemment manifesté les signes d'une renaissance[15]. Les défenseurs de cette nouvelle version de la thèse de l'identité type/type caractérisent habituellement les identités de correspondance comme étant empiriques mais nécessaires. Elles sont entendues comme étant des identités nécessaires *a posteriori* dans

---

13. Voir Feigl, H., « The "Mental" and the "Physical" », *Minnesota Studies in the Philosophy of Science*, 2, Minneapolis, University of Minnesota Press, 1958, p. 370-497 ; et Smart, J. J. C., « Sensations and Brain Processes », *Philosophical Review*, 68, 1959, p. 141-156.

14. Je crois que Hilary Putnam fut le premier à utiliser cette expression (ou plus précisément l'expression « *theoretical identification* ») ; voir Putnam, H., « Minds and Machines », dans Hook, S., dir., *Dimensions of Mind*, New York, New York University Press, 1960.

15. Voir notamment Hill, C. et McLaughlin, B., « There are Fewer Things in Reality than Dreamt of in Chalmers' Philosophy », *Philosophy and Phenomenological Research*, 59, 1999, p. 445-454 ; Block, N. et Stalnaker, R., « Conceptual Analysis, Dualism, and the Explanatory Gap », à paraître dans *Philosophical Review*.

le sens de la thèse influente de Saul Kripke, et cela au même titre par exemple que « l'eau = H_2O » et « la chaleur = le mouvement moléculaire ».

Nous ne pouvons pas nous engager ici dans une discussion élaborée de la question de savoir si, et comment, ce modèle de réduction cerveau/esprit permet de rendre compte de l'explication réductrice[16]. Cependant, il est clair que cette forme de réduction ne fournit pas de prédictions du type de celles recherchées par les émergentistes, prédictions codifiées dans notre *desideratum* sur la réduction. Comme ces identités, qu'elles soient nécessaires ou contingentes, ne peuvent être connues qu'empiriquement, et comme elles sont requises pour toute prédiction d'occurrences mentales reposant sur de l'information relative aux neurones, par exemple pour la prédiction d'occurrences de la douleur, il ne sera pas possible de prédire de telles occurrences *uniquement* sur la base d'information neurobiologique ou behaviorale à propos des organismes. Une connaissance préalable de la corrélation entre les états de douleur et ceux du cerveau doit être supposée acquise dans toute prédiction de ce type, ce qui viole le *desideratum* sur la réduction.

Lorsqu'il est appliqué au problème de la relation corps/esprit, le modèle de la réduction fonctionnelle offre l'attrait de rendre intelligible la façon dont le mental est lié aussi bien au comportement qu'au cerveau. Selon ce modèle, le mental dépend conceptuellement du comportement dans la mesure où chaque propriété mentale peut être définie comme étant un intermédiaire causal entre les stimuli, comme inputs, et le comportement (et peut-être d'autres propriétés mentales), comme outputs[17]. Cette dépendance est donc conceptuelle et nécessaire. D'autre part, les réalisateurs de ces propriétés mentales fonctionnalisées sont des propriétés et des fonctions neuronales. Le fait qu'une propriété neuronale donnée réalise la douleur, par exemple, est quelque chose d'empirique et de contingent. Empirique pour la raison évidente que la recherche scientifique sophistiquée est requise si nous voulons identifier l'état neuronal qui joue le rôle causal assigné à la douleur ; et contingent puisque le fait qu'une propriété neuronale donnée puisse ou non jouer ce rôle causal dépend, tout comme le caractère intrinsèque de la propriété elle-même, des lois causales qui prévalent dans un monde donné, et nous supposons généralement que ces lois prévalent seulement de façon contingente. Somme toute, selon le modèle de la réduction fonctionnelle, les esprits dépendent du comportement de façon conceptuelle, et du cerveau de façon contingente.

Mais rien de tout cela n'implique que le mental soit en fait fonctionnellement réductible ou non. Selon le modèle de la réduction que je défends ici, cette question revient à la suivante : les propriétés mentales peuvent-elles

---

16. Cette question est discutée dans Kim, J., « Reduction, Reductive Explanation, and the Explanatory Gap », à paraître.

17. Ou de façon holiste via la ramseyfication ; voir Lewis, D., « Psychophysical and Theoretical Identifications », dans Block, N., dir., *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980.

être fonctionnalisées? Là est la question cruciale. La raison en est qu'à partir du moment ou nous acceptons le postulat physicaliste selon lequel les propriétés mentales n'ont pas de réalisateurs non physiques, une propriété mentale fonctionnalisée doit avoir des réalisateurs physiques, et ceux-ci devraient pouvoir être découverts, en principe, grâce à la recherche scientifique. Et, puisqu'une propriété mentale est exemplifiée en une occasion particulière, nous savons qu'un de ses réalisateurs physiques doit être réalisé à cette occasion. De sorte que si une propriété mentale a été fonctionnalisée, le reste de la tâche revient à la science, et d'un point de vue métaphysique, c'est comme si elle avait été réduite. Si nous savons qu'une propriété mentale est en principe fonctionnalisable, nous savons qu'elle est réductible.

Cela signifie que le fonctionnalisme classique du type de celui qui a été formulé par Armstrong et Lewis est correct. Le mental serait physiquement réductible, puisqu'il serait fonctionnalisable. Si c'est le cas, est-il réductible au cerveau, comme l'ont préconisé Armstrong et Lewis? Cela dépend : pour ce qui est des propriétés mentales des humains il est pratiquement certain que leurs réalisateurs sont des propriétés du cerveau. Et la même chose doit être vraie de toutes les créatures terrestres douées de mentalité, c'est-à-dire ayant des propriétés mentales. Mais ce n'est qu'un début. Comme nous l'avons vu, la réduction devra procéder cas par cas, et elle procédera vraisemblablement en se concentrant sur l'identification des réalisateurs neuronaux du mental chez les humains et chez des espèces animales proches de l'espèce humaine (ce qui, comme j'en suis sûr, est ce que la recherche scientifique sur la conscience fait présentement). Il ne sera pas possible au cours de l'existence de l'espèce humaine d'identifier tous les réalisateurs actuels et possibles du mental, et en ce sens la réduction corps/esprit ne peut jamais être complétée. Nous n'avons peut-être même pas un intérêt sérieux, à long terme, à aller beaucoup plus loin que les être humains et quelques espèces animales qui nous offrent un intérêt particulier. Si nous réussissons à identifier, chez les humains, les réalisateurs neuronaux/physiques de toutes les propriétés mentales que les humains sont capables d'exemplifier, nous aurons alors acquis une compréhension réductrice complète de la mentalité humaine. Rien de plus n'est requis. Il me semble que nos succès dans l'identification, même approximative et sommaire, des réalisateurs neuronaux de certaines propriétés et fonctions sélectionnées chez les humains, suffiraient à faire valoir le réductionnisme corps/esprit, dès lors qu'ils sont combinés à la thèse métaphysique générale du physicalisme selon laquelle les propriétés mentales n'ont de réalisateurs que physiques.

Mais ces commentaires ne s'appliquent qu'aux seules propriétés mentales qui sont fonctionnalisables. Notre dernière question est donc la suivante : est-ce que toutes les propriétés mentales sont fonctionnalisables? Comme nous le savons, cette question a suscité de vives controverses, et nous ne pouvons ici nous engager dans une revue et une discussion des divers problèmes qu'elle soulève. Je me contenterai de signaler mon vote négatif, et cela

sans arguments à l'appui. Il y a en effet de bonnes raisons de croire que les propriétés mentales intentionnelles, comme la croyance et le désir, sont effectivement fonctionnalisables, mais que les caractères qualitatifs de nos expériences, c'est-à-dire ce qu'on a appelé les *qualia*, ne le sont pas. Pour cela, nous n'avons pas besoin d'endosser l'hypothèse spéculative très controversée des « zombies », ces créatures qui sont nos répliques fonctionnelles mais qui seraient totalement dépourvues de conscience phénoménale[18]. Tout ce dont nous avons besoin, c'est de la possibilité du spectre inversé. Cette possibilité porte également à controverse, mais elle est manifestement moins problématique que celle des zombies.

Les émergentistes classiques de la première moitié du XX^{e} siècle ont soutenu que plusieurs aspects significatifs d'un tout complexe sont émergents, en ce sens que leur occurrence ne peut être ni prédite ni expliquée dans les termes des conditions sous-jacentes dont elles émergent. S'il doit y avoir au moins une chose à propos de laquelle ils ont vu juste, il semble bien que leur meilleure chance se trouve du côté des qualités phénoménales de l'expérience consciente. Que ces propriétés soient émergentes ou, sinon, réductibles à des traits biologiques ou physiques des organismes, cela dépendra, comme je l'ai soutenu, de la réponse à la question de savoir si elles peuvent ou non être expliquées comme étant des propriétés fonctionnelles définies par leurs rôles causaux. Cela est donc la question cruciale à laquelle nous devons répondre, afin d'être en mesure de répondre à la question plus générale de la réductibilité physique du mental. J'espère que, si nous n'avons pas réussi à répondre à cette dernière question de façon définitive, nous avons au moins réussi à bien la préciser[19].

---

18. Pour une discussion plus détaillée de l'hypothèse des « zombies », voir Chalmers, D., *The Conscious Mind*, Oxford, Oxford University Press, 1996.

19. Ce texte a été traduit de l'anglais par Richard Vallée et Paul Bernier.