

Naturalisme émergentiste et explication causale

Jean-Michel ROY*

RÉSUMÉ. Le problème de l'exclusion de l'efficacité causale des propriétés mentales par les propriétés physiques a été formulé essentiellement en référence au matérialisme non réductionniste de type fonctionnaliste. La question se pose donc de savoir si cette autre forme de matérialisme non réductionniste que prétend le plus souvent être le naturalisme émergentiste permet de le surmonter. Ce qui implique tout d'abord de se demander si la différence entre propriété fonctionnelle et propriété émergente est telle que le problème de l'exclusion ne se pose pas même pour l'émergentisme. L'article répond par la négative en défendant l'idée que ce problème est lié au concept de propriété abstraite et irréductible, et que la différence entre propriété fonctionnelle et propriété émergente recouvre une différence d'espèce au sein du genre de la propriété abstraite et irréductible.

Mots clefs : problème de l'exclusion, matérialisme non-réductionniste, propriété abstraite, fonctionnalisme, émergentisme.

ABSTRACT. Naturalist emergentism and causal explanation. The problem of the exclusion of the causal efficacy of mental properties by physical properties has been mainly raised in connection with non reductionist materialism of a functionalist kind. It is therefore legitimate to wonder whether the alternative form of non reductionist materialism that emergentist naturalism often pretends to incarnate can overcome the exclusion problem. And it is first necessary to examine whether the difference that separates the notions of functional and emergentist properties is such that naturalist emergentism does not face the exclusion problem. The article answers negatively to this question by arguing that the exclusion problem is rooted in the notion of an abstract and irreducible property, and that the difference between functional and emergentist properties amounts to one between two species of such an abstract and irreducible property.

Key words: exclusion problem, non reductionist materialism, functionalism, abstract property, emergentism.

1. DU FONCTIONNALISME À L'ÉMERGENTISME¹ :

Un des problèmes essentiels de la réflexion philosophique contemporaine sur la possibilité de construire une théorie scientifique des phénomènes cognitifs est à l'évidence celui de la naturalisation des propriétés mentales. Et la doctrine dite du matérialisme non-réductionniste, qui a pris différentes figures mais dont la plus connue est celle du fonctionnalisme, notamment sous ses formes computationnelle et causale, constitue sans aucun doute l'un de ses apports majeurs à cet égard. Le fonctionnalisme se heurte néanmoins à un

* École Normale Supérieure de Lettres et Sciences Humaines, e-mail : roy@heraclite.ens.fr.

¹ Ce texte est issu d'une intervention au colloque *Du constructivisme aux lois de la pensée* organisé en 1998 par M.J. Durand-Richard. Son ultime révision a bénéficié du soutien du programme de recherches CPER *Représentationalisme et Fondements des Sciences Cognitives*, conduit par l'ENS-LSH avec le concours de la Région Rhône-Alpes.

certain nombre de difficultés désormais bien établies ; ce qui ne signifie en aucune façon que l'on peut les considérer pour autant comme bien connues, et encore moins comme l'ayant invalidé de façon définitive. On peut craindre au contraire que des théories qui se présentent comme des alternatives par rapport à lui ne font souvent, ou que reproduire subrepticement certaines des difficultés qui lui sont imputées faute d'avoir pris le temps de les examiner avec suffisamment de soin et d'objectivité, ou bien lui adresser pour les mêmes raisons des reproches purement dogmatiques.

C'est dans cet esprit que je souhaite me pencher ici sur la question de savoir si une conception émergentiste des propriétés mentales, qui se trouve assez étroitement liée avec ce que recouvre le terme de constructivisme épistémologique, peut permettre de résoudre l'un des plus graves obstacles auxquels se heurte une interprétation réaliste du fonctionnalisme, en vertu de laquelle les propriétés mentales sont des propriétés fonctionnelles dotées d'efficacité causale et permettant de construire de véritables explications causales des comportements cognitifs. Cet obstacle consiste en ce que les propriétés physiques dont dépendent les propriétés mentales ainsi conçues semblent devoir nécessairement exercer à leur égard une sorte de droit de préemption en matière d'efficacité causale, et par là leur dérober tout pouvoir causal; et ce, alors même que, dans une perspective réaliste, le fonctionnalisme est précisément conçu comme une façon de parvenir à enfin donner une forme scientifiquement satisfaisante à la vieille idée de causalité mentale, et à celle de causalité psycho-physique en particulier. La littérature contemporaine désigne alternativement cette difficulté du nom de problème de la préemption (de l'efficacité causale des propriétés physiques par rapport aux propriétés mentales), ou de celui de problème de l'exclusion (de l'efficacité causale des propriétés mentales par celle des propriétés physiques)². La question que je souhaite examiner est donc très précisément la suivante : est-ce que la substitution de l'idée qu'une propriété mentale est une propriété émergente de propriétés physiques à celle qu'elle est une propriété fonctionnelle de propriétés physiques permet d'éliminer un tel obstacle et de faire tomber l'accusation d'épiphénoménisme à laquelle il conduit? En d'autres termes : l'émergentisme peut-il représenter une forme de réalisme mental matérialiste non-réductionniste plus satisfaisante à cet égard que le fonctionnalisme ?

La façon la plus simple et la plus radicale pour qu'il en aille ainsi serait à l'évidence que la notion de propriété émergente, à la différence de celle de propriété fonctionnelle, ne soulève pas le problème de l'exclusion, ou le soulève au pire d'une façon qui en rend la solution transparente et immédiate. Et c'est la seule question de savoir si tel est effectivement le cas que les lignes qui suivent se proposent d'examiner. Elles font ainsi valoir quelques raisons de penser que fonctionnalisme et émergentisme entretiennent des relations de cousinage conceptuel trop étroites pour ne pas être parfaitement égaux devant la difficulté, mais aussi et de façon plus profonde, que le problème de l'exclusion s'enracine dans ce qui leur est commun, à savoir l'idée générale de matérialisme non-réductionniste.

Il importe de bien souligner le caractère limité des conséquences de la réponse affirmative qui se trouve avancée. Celle-ci signifie seulement que

² La littérature sur ce point est trop importante pour pouvoir être indiquée intégralement. On se contentera de renvoyer à J. Kim (1993), qui recueille des articles séminaux à cet égard, et à McDonald C. & G. (1995), qui contient quelques contributions importantes du débat.

l'émergentisme ne peut se soustraire à l'exigence de résoudre le problème de l'exclusion, notamment par ignorance ou indifférence. Mais elle n'implique en aucune manière qu'il ne dispose pas des armes nécessaires pour en venir à bout, pas plus d'ailleurs que l'hypothèse contraire. Pour en décider, il faut s'assurer au préalable qu'il se heurte effectivement à lui. Elle suggère par contre que ces armes ne sont ni supérieures ni inférieures en nombre et en qualité à celles qui sont entre les mains du fonctionnalisme.

Je commencerai par faire retour de manière circonstanciée sur la notion de matérialisme non-réductionniste. Approfondissant ensuite l'analyse de ce qui, dans le fonctionnalisme, est à la source du problème de l'exclusion, j'arguerai que la responsabilité en incombe à la notion de propriété abstraite, et que celle-ci étant une détermination essentielle du concept de matérialisme non-réductionniste, le problème de l'exclusion se reporte automatiquement du fonctionnalisme sur l'émergentisme. Cette hypothèse sera alors vérifiée en faisant successivement apparaître qu'une propriété émergente est bien une propriété abstraite, et que rien de ce qui fait la spécificité de l'abstraction de type émergentiste ne paraît susceptible de pouvoir bloquer les implications épiphénoménistes attachées à sa dimension générique.

2. MATÉRIALISMES RÉDUCTIONNISTE ET NON-RÉDUCTIONNISTE:

Comment convient-il en premier lieu d'entendre la notion de matérialisme non-réductionniste ? La définition que l'on décide d'en retenir est à l'évidence d'une importance cruciale pour les différents aspects de la question soulevée, puisque c'est d'elle que dépend la légitimité aussi bien du problème de la préemption de la causalité physique que de la solution que l'on peut en proposer, et tout particulièrement de celle que l'on peut éventuellement attribuer à l'émergentisme. Or la structure même de l'expression de matérialisme non-réductionniste indique assez clairement que son contenu doit être appréhendé à partir du concept de matérialisme réductionniste. Ce qui contraint donc à commencer par fournir une caractérisation relativement précise de celui-ci.

Ce double travail de définition est d'autant plus indispensable que règne dans la littérature actuelle un certain flou terminologique, et que par ailleurs tant l'expression de matérialisme réductionniste que celle de matérialisme non-réductionniste sont dans une certaine mesure trompeuses. D'une part on n'opère le plus souvent aucune distinction entre cette dernière et celle de fonctionnalisme, alors même que le fonctionnalisme n'est pas par lui-même matérialiste, et qu'à l'inverse tout matérialisme non-réductionniste n'est pas de type fonctionnaliste. D'autre part on peut parfaitement défendre, ainsi qu'on le fera valoir plus loin, que, au moins sous l'une de ses formes et selon l'une des définitions possibles du réductionnisme, le matérialisme dit non-réductionniste mérite parfaitement la qualification de réductionniste, et que ce qui se trouve ordinairement désigné par cette dernière doit par conséquent être conçu comme une simple espèce et non pas comme le genre. Il convient donc de spécifier minutieusement chacune des notions mises en jeu, sous peine non seulement de ne pouvoir fournir une formulation suffisamment précise du problème de la préemption, mais aussi de voir inutilement contester les prémisses sur lesquelles il repose.

21. Matérialisme et physicalisme:

De toutes ces notions, la plus fondamentale est très certainement celle de matérialisme elle-même. Le terme de physicalisme est en fait plus adéquat,

bien qu'il ait été par le passé le plus souvent réservé à une forme spécifiquement réductionniste de matérialisme. Aussi dira-t-on ici qu'une théorie est physicaliste, ou encore est un physicalisme, dès lors qu'elle n'admet dans son ontologie (au sens de ce à quoi elle accorde la réalité) que des entités physiques. Celles-ci se définissent à leur tour comme l'ensemble des entités postulées par les sciences physiques à un moment donné de leur développement, par opposition à ce qui satisfait une définition philosophique de ce qui est physique, même si une telle opposition ne saurait être conçue de façon trop tranchée.

Ainsi définie, la notion de physicalisme ne recouvre pas celle de matérialisme, puisque d'une part elle admet l'existence d'autres entités que la matière, et que d'autre part elle exclut les conceptions de la matière qui diffèrent de celle qui est à l'œuvre dans les sciences physiques à un état E de leur évolution. Considérer que tout ce qui est est une entité physique *ainsi définie* ne revient cependant pas à affirmer que tout ce qui est n'est rien d'autre qu'une entité physique telle que la caractérisent les sciences physiques à l'état E, mais simplement que rien n'est sans être une entité physique telle que la caractérisent les sciences physiques à l'état E. C'est, en d'autres termes, simplement épouser ce que Jerry Fodor appelle avec justesse la thèse de la généralité de la physique : tout ce qui est appartient au domaine d'investigation des sciences physiques. Une telle notion de physicalisme est par ailleurs plus étroite que celle de naturalisme, entendue comme l'affirmation que tout ce qui est appartient au domaine d'investigation des sciences de la nature. Elle n'est pas en opposition avec elle toutefois, dès lors que l'on admet que tout ce qui appartient à ce dernier appartient à celui des sciences physiques, ainsi promues au titre d'élément fondamental des sciences de la nature.

Le concept de physicalisation, ou encore de naturalisation, peut alors se définir essentiellement, du point de vue ontologique, comme l'explication rationnelle du fait qu'une propriété, en apparence hétérogène avec les propriétés caractérisent les entités physiques comme telles, appartient néanmoins *proprio sensu* à ces dernières, et est donc du fait même compatible avec leurs propriétés caractéristiques. Ce qui exige de montrer qu'elle leur appartient *en vertu même* de ces propriétés qui les caractérisent en tant qu'entités physiques, et du même coup qu'elle est ontologiquement dépendante par rapport à elles. Par le biais d'une telle explication, la propriété en question est transformée en propriété physique, et se trouve donc physicalisée ou naturalisée.

L'idée d'appartenance *proprio sensu* s'oppose à celle d'une appartenance qui ne pourrait être établie que par le biais de la possession d'une propriété non-physique. Car en un tel cas, la propriété n'appartiendrait pas aux entités physiques *en tant que telles*, et ne serait donc pas elle-même physicalisée. L'exigence d'explication s'oppose quant à elle au fait de simplement traiter la propriété apparemment hétérogène aux propriétés physiques comme l'une d'elles, en se contentant de l'attribuer *de facto* à des entités physiques. Une telle attitude élimine le problème de la physicalisation sans le résoudre, mais sans montrer non plus son illégitimité. Toute la difficulté est en effet de comprendre comment une entité physique peut posséder une propriété qui n'entre pas dans sa caractérisation en tant qu'entité physique à un état E du développement des sciences physiques. Dès lors que l'on décide de traiter une propriété apparemment hétérogène aux propriétés caractéristiques des entités physiques comme une propriété physique, et qu'on a par là refusé de s'engager dans la voie d'un dualisme des substances, cette difficulté exige une réponse, sauf à substituer à

un dualisme des substances un dualisme des propriétés qui risque de n'en être qu'une version masquée³.

Dans le contexte du débat contemporain sur le problème de la préemption, et plus généralement de la naturalisation du mental, la différence entre les termes de physicalisme et de matérialisme qui vient d'être soulignée n'est néanmoins pas respectée. Aussi le second désigne-t-il en fait pour l'essentiel une telle conception physicalisée des propriétés mentales⁴. C'est du moins en ce sens qu'il sera seul pris ici.

22. Le concept de matérialisme réductionniste :

Le matérialisme réductionniste apparaît alors à son tour comme une forme particulière de physicalisation des propriétés mentales, dont la spécificité consiste à affirmer que celle-ci peut être obtenue au moyen d'une opération de réduction.

L'expression de réduction, qui est empruntée au langage ordinaire, a connu bien des usages philosophiques. Il semble pourtant que, dans la quasi totalité de ses emplois techniques, son sens conserve une détermination générale commune, et qui est celle de l'annulation d'une différence entre deux termes. On réduit en effet toujours une chose à une autre (en accordant au terme de chose une acception suffisamment large), et ce faisant on supprime la différence qui les sépare, de façon à en obtenir une seule là où on en comptait deux.

Ainsi entendue, la physicalisation réductionniste des propriétés mentales consiste donc à annuler la spécificité de ces dernières en les ramenant à des propriétés dégagées par la physique. Mais les termes de la réduction qu'elle met en œuvre sont en réalité plus larges : celle-ci est en effet inter-théorique et consiste donc à annuler la différence qui sépare une théorie mentaliste et une théorie physique, chacune d'elles étant à son niveau le plus général conçu comme un système déductif. Cette différence réside certes fondamentalement dans le fait que les propositions de la première portent sur des propriétés mentales et celles de la seconde sur des propriétés physiques. Toutefois, réduire une science à une autre exige plus que de réduire les termes désignant les propriétés qui lui sont propres à des termes désignant celles de l'autre. Il faut encore, ainsi qu'y ont insisté tous les théoriciens de la réduction interthéorique de Carnap à Quine, que les propositions de la première science ainsi transformées en propositions de la seconde soient dérivées déductivement de celles qui sont originellement constitutives de celle-ci. L'annulation de la différence de nature entre deux propriétés peut être obtenue de deux plusieurs façons. Elle peut être opérée par l'élimination pure et simple de l'un des deux termes en rapport, donnant alors naissance à ce que l'on appelle communément l'éliminativisme, ou encore la théorie de la disparition⁵. Mais c'est là au fond un cas limite de réductionnisme, puisque l'on ne réduit pas véritablement la différence, on nie d'emblée son existence. Il est certes tentant d'objecter que réduire une différence revient précisément à la supprimer, et à donc à éliminer le terme réduit au profit du terme réducteur. C'est pourtant une chose que

³ Le dualisme des propriétés consiste donc à attribuer les propriétés mentales à des entités dotées de propriétés physiques, mais à les considérer comme inexplicables rationnellement à partir de ces dernières, et partant, comme des propriétés d'un autre type ou catégorie qu'elles.

⁴ Il conviendrait ainsi notamment de parler de physicalisme non-réductionniste plutôt que de matérialisme non-réductionniste.

⁵ L'expression est due à Richard Rorty (cf. Rorty, 1970).

d'affirmer que rien de tel qu'une propriété mentale n'existe, et une autre que de soutenir qu'une propriété mentale n'est rien d'autre qu'une propriété physique, de la même façon que c'est une chose d'affirmer qu'il n'y a rien de tel que les fantômes, et une autre que de faire valoir qu'ils ne sont rien d'autre que des phénomènes physico-physiologiques exceptionnels. Dans le premier cas on rejette la réalité des propriétés mentales, dans le second on rejette le caractère distinct, par rapport aux propriétés physiques, des propriétés mentales.

Aussi seule cette seconde manière d'annuler la différence du mental et du physique constitue-t-elle à proprement parler une réduction, faisant du même coup apparaître que celle-ci est fondamentalement une opération d'identification. Cette identité est de plus établie à un niveau spécifique, de telle manière que chaque *type* de propriété mentale se trouve identifié à un *type* déterminé de propriété physique; d'où le nom de physicalisme des types donné au physicalisme réductionniste. Il convient toutefois souligner que l'identification d'une chose à une autre peut être obtenue soit en négligeant, soit en niant les différences qu'elle présente par rapport à elle, même si toutes deux débouchent sur la substitution de la seconde à la première. Dans le premier cas, on décide simplement de ne pas tenir compte de différences dont la réalité n'est pas contestée, tandis que dans le second cette réalité est contestée. La première démarche n'est cependant véritablement légitime que pour autant que l'on mette par ailleurs à nu des identités. C'est ainsi par exemple que, dans son ouvrage de 1928 qui représente sa première tentative pour réunir en un vaste système réductionniste l'ensemble des concepts scientifiques⁶, Carnap définit la réduction ontologique comme une identification fondée sur la coextensionnalité des fonctions propositionnelles. Mais dire que deux fonctions propositionnelles peuvent être substituées l'une à l'autre parce qu'elles sont vraies des mêmes objets n'est pas affirmer encore qu'elles sont identiques, mais simplement qu'elles peuvent être traitées comme telles. Et de fait, Carnap précise qu'une telle méthode de réduction ne présuppose nullement que les fonctions concernées soient intensionnellement identiques, mais que cette différence intensionnelle est seulement négligée.

Le réductionnisme ainsi défini comme l'identification des différents types de propriétés mentales à différents types de propriétés physiques dans le but de transformer la connaissance scientifique du mental en une partie des sciences physiques est encore susceptible de recevoir des formes assez diverses. Les oppositions qui séparent les deux principales réalisations qu'il a connues à l'époque contemporaine en fournissent une bonne illustration. Les identifications que propose chacune de ces dernières diffèrent en effet sur plusieurs points :

a) *le contenu* : Tandis que le physicalisme de Carnap et de Hempel, qui constitue la figure de proue du positivisme logique en la matière, identifie la propriété mentale principalement avec une disposition comportementale⁷, et reçoit pour cette raison même le nom de béhaviorisme logique, la théorie de l'identité centrale représentée par U.T. Place, J.J.C. Smart ou D.M. Armstrong⁸ l'assimilent à un état du système nerveux central. Si aucune de ces propriétés réductrices n'est elle-même une propriété physique *stricto sensu*, elles sont néanmoins considérées comme réductibles à leur tour à de telles propriétés.

⁶ Carnap (1928).

¹¹ Cf. par exemple Carnap (1932-33).

⁸ Les articles emblématiques de ces trois auteurs sont rassemblés dans Borst (1970).

b) *la problématique* : Tandis que le béhaviorisme logique de Carnap et Hempel se donne comme une thèse relative à ce que signifient les termes mentaux, tant dans leur usage ordinaire que scientifique, la théorie de l'identité centrale conçoit au contraire ses identifications comme des thèses relatives à la nature même de ce qui se trouve être signifié. Pour subtile qu'elle soit, la différence est importante. Ainsi U.T. Place écrit-il dans « Is consciousness a brain process ? », article *princeps* de la théorie de l'identité centrale : « en défendant la thèse que la conscience est un processus du cerveau, je ne prétends pas que quand nous décrivons nos rêves, nos fantaisies et nos sensations, nous parlons de processus cérébraux »⁹. C'est une chose en effet de dire que le terme de conscience ne signifie rien d'autre que ce que signifie l'expression de processus cérébral de tel ou tel type, et une autre que d'affirmer que ce que nous signifions au moyen du terme de conscience est en réalité un processus cérébral. La distinction en jeu est au fond celle, traditionnelle, de la *definitio nominis* et de la *definitio rei*.

c) *statut épistémologique* : Cette différence de problématique a une conséquence directe sur le statut épistémologique des deux types d'identifications. Tandis que le béhaviorisme logique les conçoit comme des vérités analytiques et nécessaires, la théorie centrale de l'identité y voit au contraire des énoncés synthétiques et contingents.

23. La définition du matérialisme non-réductionniste:

C'est en très large mesure en réaction contre ces deux formes précises de matérialisme ou de physicalisme réductionniste que s'est historiquement constitué le matérialisme non-réductionniste, ainsi qu'en témoignent avec une parfaite clarté « La nature des états mentaux » de H. Putnam¹⁰, qui en est l'un des principaux manifestes. Ou encore les analyses légèrement postérieures exposées par Fodor en 1968 dans son ouvrage *Psychological Explanation*¹¹, puis dans un article de 1974 rebaptisé un an plus tard du titre révélateur de "Two kinds of reductionism" pour servir d'introduction à son *Language of Thought*¹². Il est vrai toutefois que, dans la mesure où la théorie de l'identité centrale est elle-même en partie le produit d'une dénonciation des insuffisances du béhaviorisme logique, c'est principalement en opposition avec cette dernière que le matérialisme non-réductionniste a défini ses thèses essentielles. Il convient en outre de préciser que cette réaction contre la version neurobiologique du physicalisme réductionniste a pris d'autres formes que le matérialisme non-réductionniste, et notamment celles d'un retour au dualisme¹³ ou encore d'un éliminativisme¹⁴.

Or que rejette le matérialisme non-réductionniste dans ces deux doctrines, et notamment dans celle défendue par la théorie de l'identité centrale? Même face à la nécessité de s'en tenir à ses figures originelles et principales pour tenter de répondre à cette question, on ne saurait bien entendu limiter le matérialisme non-réductionniste aux formes que lui ont données H. Putnam et J. Fodor. Des théories comme celles de Donald Davidson, baptisée monisme anomal et principalement défendue en 1970 dans « Les événements mentaux »

⁹ Borst (1970, p. 44).

¹⁰ Cf. Putnam (1967).

¹¹ Putnam (1968).

¹² Fodor (1975).

¹³ Popper (1977).

¹⁴ Quine (1960).

(Davidson 1980¹⁵), ou de David Lewis, présentée en 1966 dans “An argument for identity theory”¹⁶ et précisée par la suite dans plusieurs articles¹⁷, doivent impérativement être prises en compte. Quels sont donc les points fondamentaux sur lesquels se rejoignent leurs critiques et qui leur valent de venir se ranger sous une bannière commune ?

231. Les thèses négatives :

Ils se résument pour l'essentiel à une dénonciation du caractère trop étroit des identifications proposées par les deux figures majeures du matérialisme réductionniste, au prétexte qu'il serait en réalité impossible d'établir les corrélations invariables entre *types* d'états mentaux et *types* de dispositions comportementales ou types d'états du système nerveux qu'elle requiert. Quand elle est dirigée contre la théorie de l'identité centrale, cette analyse prend la forme plus particulière de l'argument dit de la réalisabilité multiple, qui fait valoir qu'un même état mental peut être attribué à deux entités physiques sans qu'il soit pour autant possible de considérer que celles-ci soient dans le même état physique. Cet argument connaît en réalité plusieurs versions, différant l'une de l'autre par leur degré de précision et de sophistication, ainsi que par les motifs qu'elles invoquent.

La plus simple, en même temps que la plus ancienne sans doute, est celle présentée par H. Putnam dans « La nature des états mentaux ». Celui-ci se contente en effet d'affirmer qu'il est éminemment probable qu'un type déterminé d'état mental, telle que la faim ou la douleur, soit *de facto* corrélée parmi les espèces d'organismes capables d'en être dotés avec des états physico-chimiques de leurs cerveaux de types différents, et qu'une seule exception de ce genre suffit à ruiner la thèse générale de la théorie de l'identité centrale. Putnam s'attache donc très précisément à remettre en cause, non pas la possibilité de conclure de l'existence d'une corrélation invariable entre la possession d'un type d'état mental et celle d'un type d'état cérébral à l'identité de ces deux derniers, mais la possibilité même d'établir une telle corrélation.

La version de l'argument de la réalisabilité multiple avancée par Fodor dans l'introduction à son ouvrage de 1975 dégage avec plus de précision son point essentiel, et qui est bien incontestablement aussi celui que cherche à faire valoir Putnam. Fodor introduit en effet un *distinguo* entre le physicalisme des types et le physicalisme réductionniste proprement dit, bien qu'il semble en fait en certains endroits utiliser l'expression de physicalisme des types pour désigner aussi la seconde de ces deux choses. Envisagés dans leur dimension la plus générale, le premier se contente de supposer que chaque type de propriété non-physique peut être corrélé, et donc identifié, avec un type de propriété physique, tandis que le second fait l'hypothèse que chaque type de propriété non-physique *nomique*, c'est à dire susceptible de permettre la formulation d'une loi ou encore, selon la terminologie adoptée par Fodor, de déterminer une espèce, peut être corrélé, et donc identifié, avec une propriété physique *nomique*. Il écrit ainsi : « Si le réductionnisme est vrai, chaque espèce est, ou est coextensive avec, une espèce physique »¹⁸. Or seul ce second type de

¹⁵ Cf. également "Psychology as Philosophy" (Davidson, 1974) et "The Material Mind" (Davidson, 1973).

¹⁶ Cf. Lewis 1966.

¹⁷ Notamment : "How to define theoretical terms" (Lewis, 1970), et "Psychophysical and theoretical identifications" (Block, 1980).

¹⁸ Fodor (1975, p. 14).

corrélation est à ses yeux improbable dans le cas de l'identification du mental et du cérébral. Car il soutient de manière convaincante que le physicalisme des types *stricto sensu* ne se trouve pas remis en cause par le simple fait qu'un type déterminé de propriété mentale peut être corrélé avec plusieurs types différents de propriétés cérébrales : il suffit en effet alors de considérer que cette propriété mentale est en fait corrélée avec un type *complexe* de propriété cérébrale, constituée par la conjonction de différentes propriétés cérébrales simples, ou plus simples. Aussi le problème est-il bien plutôt que cette propriété complexe a fort peu de chances de permettre la formulation d'une loi neurobiologique. Les arguments avancés en faveur d'une telle improbabilité demeurent cependant essentiellement indirects et assez peu développés. Fodor se contente en effet d'en fournir une illustration particulière dans le cas de l'économie, et d'affirmer qu'elle peut être aisément transposée à celui de la psychologie mentaliste, sans analyser plus avant ni le bien-fondé de cette illustration, ni les raisons qui permettent de croire qu'elle a valeur d'exemple pour elle. Ainsi fait-il valoir que la propriété d'être un d'échange monétaire, à laquelle fait appel la loi de Gresham, peut-être considérée comme corrélée avec celles d'être un échange de billets verts, d'être un échange de pièces d'or, d'être un échange d'écus...., sans que la propriété complexe formée de la disjonction de ces dernières soit un candidat probable à la formulation d'une loi physique, au sens où il est n'est guère plus probable que des choses physiques soient soumises à une loi en vertu du fait de posséder cette propriété que d'autres ne le soient en vertu du fait de posséder la propriété d'être situées à telle ou telle distance de la Tour Eiffel. Car Fodor admet que chacun des différents types de propriétés physiques avec lesquels peut être corrélé un même type de propriété mentale peuvent être eux-mêmes nomiques. Mais cette possibilité ne change rien à l'obstacle qui se dresse devant le réductionnisme; car à ses yeux le fait d'être une disjonction de propriétés nomiques n'implique en aucune façon celui d'être une propriété nomique.

Il est vrai que Fodor développe quelques arguments spécifiques à propos de la psychologie; mais curieusement ceux-ci ne concernent pas le point essentiel de sa critique, à savoir l'improbabilité que la propriété cérébrale complexe qui peut être corrélée avec une propriété mentale soit elle-même nomique. Ils portent en effet d'une part sur le fait que les corrélations demeurent jusqu'à maintenant très approximatives (ce qui est un point logiquement antérieur au fait que la propriété cérébrale corrélée ait peu de chances d'être elle-même nomique), et d'autre part sur le fait qu'elles ont fort peu de chance d'être elles-mêmes nomiques. En effet, Fodor distingue la nomicité de la propriété avec laquelle se trouve corrélée la propriété mentale de celle de la corrélation elle-même. Il écrit ainsi : « En fait la situation du réductionnisme est pire encore. Car le réductionnisme affirme non seulement que toutes les espèces sont co-extensives avec les espèces physiques, mais que ces co-extensions sont nomologiques: les lois ponts sont des lois »¹⁹. C'est d'ailleurs aussi le point essentiel de la critique davidsonienne. Davidson admet en effet sans réserve dans « Les événements mentaux » qu'il est toujours possible de trouver un type de propriété cérébrale correspondant à chaque type de propriété mentale, car celle-ci n'est rien d'autre que la disjonction des différentes propriétés physiques auxquelles la propriété mentale se trouve associée dans ses diverses

¹⁹ Fodor (1975, p. 20).

réalisations matérielles²⁰. De plus, en vertu de la solution qu'il adopte au problème de la causalité mentale, il admet également que chacun des composants de cette propriété complexe est, ou du moins peut être lui-même une propriété nomique. Sa critique diffère principalement de celle de Fodor au niveau de l'analyse de ce qui rend impossible la corrélation de la propriété mentale avec une propriété cérébrale complexe nomique. Cette impossibilité tient moins au fait qu'aucune loi physique ne pourrait être formulée au moyen d'une telle propriété, qu'à celui qu'une telle corrélation ne présente pas les caractères d'une loi. Aussi sa thèse critique principale est-elle ainsi formulée: « ...aucun prédicat purement physique, aussi complexe soit-il, n'a, en vertu d'une loi, la même extension qu'un prédicat mental »²¹.

232. Les thèses positives :

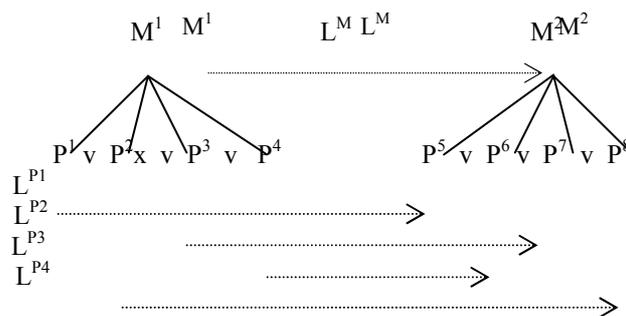
Ce rapide examen des principales critiques formulées par les chefs de file du matérialisme non-réductionniste permet donc tout à la fois de préciser la nature du matérialisme réductionniste contre lequel il se définit et de déterminer avec exactitude son point de rupture avec lui. Ce que rejettent en effet les matérialistes non-réductionnistes est la possibilité de fait ou de droit d'établir, entre les propriétés mentales nomiques auxquelles recourt une science mentaliste et des propriétés physiques nomiques, des corrélations permettant de fonder l'identification des premières avec les secondes.

Ce refus n'entraîne pas pour autant à leurs yeux la nécessité de rejeter le matérialisme ou le physicalisme tel qu'il a été défini plus haut, c'est à dire la thèse de la généralité de la physique selon laquelle tout ce qui est est une entité physique, au sens d'une entité qui possède des propriétés physiques telles que les détermine la science physique. Mais le problème est alors de trouver une explication, alternative à celle du réductionnisme, au fait que des propriétés apparemment hétérogènes avec ces dernières puissent appartenir à une entité physique. Deux solutions principales sont avancées par les critiques du matérialisme réductionniste qui définissent deux formes différentes de matérialisme non-réductionniste, qu'il importe d'autant plus de distinguer qu'elle sont été toutes deux englobées sous une même appellation de physicalisme des particuliers (*token physicalism*), par opposition au physicalisme des types.

Admettant la conclusion de l'argument de la réalisabilité multiple selon laquelle un type de propriété mentale est, au moins dans certains cas, très probablement corrélé avec une pluralité de types de propriétés physiques nomiques, l'une propose d'identifier tout simplement le premier à la disjonction des seconds, se résignant au fait que ni cette disjonction ni sa corrélation avec le type de propriété mentale ne sont nomiques. En bref, elle se contente de soustraire du concept de matérialisme réductionniste l'élément qui fait problème, renonçant du même coup à l'idée qu'une science mentaliste puisse être réduite, au sens strict du terme, à la science physique. Dans l'introduction à *Language of Thought*, Fodor illustre cette définition au moyen d'un schéma qui en résume parfaitement la thèse essentielle et que l'on peut reproduire sous une forme simplifiée de la manière suivante :

²⁰ « ...il se peut qu'il existe une phrase ouverte physique coextensive avec chaque prédicat mental », (op. cit. tr. fr. p. 289).

²¹ op. cit. tr. fr. p. 288.



La propriété M^1 est identique à la disjonction « $P^1 \vee P^2 \vee P^3 \vee \dots \vee P^x$ », tandis que la propriété M^2 est elle identique à la disjonction « $P^5 \vee P^6 \vee P^7 \dots \vee P^x$ ». Chacun des éléments des deux disjonctions est une propriété nomique, de telle sorte que la loi mentale (L^M) peut être identifiée à une disjonction de lois physiques ($L^{P^1}, L^{P^2} \dots$), et non plus à une seule comme l'exigeait le réductionnisme.

Fodor qualifie une telle solution de physicalisme des particuliers. Mais il faut prendre garde que la notion de particulier désigne ici l'instanciation d'une propriété, et que la notion de physicalisme des particuliers signifie donc que chaque instanciation d'une propriété mentale se confond avec l'instanciation d'une propriété physique. Le fait que M^1 appartienne à une entité x est indistinguable en telle occurrence avec le fait que P^1 appartienne à x , en telle autre occurrence avec le fait que P^2 appartienne à x ... Et c'est pour cette raison même que la propriété M^1 n'est rien d'autre que la propriété disjonctive ($P^1 \vee P^2 \vee P^3 \dots$). Car comment pourrait-elle n'être rien d'autre, en chacune de ses occurrences, que $P^1, P^2, P^3 \dots$, sans être identique à la disjonction de ces dernières?

Mais c'est là aussi ce qui rend une telle appellation inadéquate, du moins en tant qu'on l'oppose au physicalisme des types, ainsi que s'y emploie Fodor dans le même texte. Car si le physicalisme des types se distingue du réductionnisme, comme il le propose lui-même, par le fait d'identifier une propriété mentale avec la disjonction non-nomique des propriétés physiques avec lesquelles elle est corrélée, le physicalisme des particuliers ainsi entendu satisfait bien une telle définition. Admettre avec lui que l'instanciation d'un type de propriété mentale est identique avec l'instanciation d'un type variable de propriété physique, au sens où les deux propriétés instanciées sont en chaque cas indistinguables, conduit nécessairement à admettre qu'un type de propriété mentale est identique à un type complexe disjonctif de propriété physique, quoique non pas nomique, ni nomiquement corrélé avec lui.

La seconde solution non-réductionniste consiste, tout en admettant qu'un même type de propriété est corrélé avec différents types de propriétés physiques, à l'identifier non pas avec la propriété complexe disjonctive que celles-ci constituent, mais avec une propriété fonctionnelle commune à chacun d'eux. Une propriété fonctionnelle se définit comme une propriété qui spécifie ce que fait, par exemple du point de vue causal, une entité, par opposition à ce qu'elle est. Cette solution reste matérialiste, non seulement en ce qu'elle fait de la propriété fonctionnelle la propriété d'une chose matérielle, mais aussi en ce qu'elle explique comment elle peut l'être. C'est en effet en vertu des propriétés physiques qu'elle possède qu'une entité possède la ou les propriétés fonctionnelles qui la caractérisent. Ces dernières se trouvent donc du même coup

reconnues être dépendantes par rapport aux premières, au sens fort où celles-ci constituent des conditions nécessaires, en même temps que suffisantes, de leur instanciation. De plus, cette dépendance nécessaire prend également la forme spécifique d'une relation de survenance, en ce qu'un même type de propriété fonctionnelle peut être corrélé avec divers types de propriétés physiques, tandis qu'au contraire un même type de propriété physique ne peut être corrélé qu'avec un seul type de propriété fonctionnelle. Enfin, en vertu de ce rapport de dépendance, la propriété mentale peut être considérée comme une propriété de propriétés physiques, et à ce titre comme une propriété physique d'ordre supérieur.

Or, dans la mesure où la propriété fonctionnelle est elle-même conçue comme étant de nature nomique, il n'y a à la réflexion aucune raison de ne pas considérer le fonctionnalisme ainsi entendu comme une nouvelle forme de matérialisme réductionniste, si l'on définit simplement celui-ci comme le fait de pouvoir identifier toute propriété mentale avec une propriété physique nomique: la propriété mentale se trouve bien en effet réduite à une propriété physique nomique, quoique d'ordre supérieur et non pas de premier ordre comme dans le matérialisme réductionniste classique. Dans cette perspective, l'appellation de matérialisme non-réductionniste n'est donc pas rigoureuse. Elle ne l'est que si on prend garde de souligner que cette propriété physique de second ordre est elle-même irréductible à des propriétés physiques de premier ordre.

Le fonctionnalisme est par ailleurs également baptisé de physicalisme des particuliers. Mais il est clair que cette dernière notion n'est pas alors prise dans le même sens que précédemment. Son emploi se justifie par le fait que chaque fois qu'une propriété fonctionnelle M est instanciée, une propriété physique P doit l'être aussi. Mais les deux propriétés instanciées ne sont pas pour autant identifiées l'une avec l'autre. C'est une chose en effet que d'admettre que chaque instanciation de M est identique avec l'instanciation d'un P, et une autre que d'admettre que ce qui à chaque fois possède ainsi la propriété M n'est rien d'autre que ce qui possède une propriété P. Le fonctionnalisme n'admet que la seconde identification, et constitue de la sorte une forme différente de physicalisme des particuliers. Son théoricien le plus clair à cet égard est certainement D. Lewis, qui met parfaitement en lumière que le fonctionnalisme est une forme de théorie psychophysique de l'identité affirmant que « chaque expérience est identique avec... un état neurochimique », mais seulement au sens où ce qui, en chaque cas, a tel type de propriété mentale est aussi ce qui a tel type, variable, de propriété physique. Aussi cette seconde solution peut-elle être également illustrée au moyen du schéma précédent, à condition de voir en M¹ et M² des propriétés fonctionnelles non-identiques aux disjonctions de propriétés qui s'inscrivent sous elles, et de la même façon, de considérer que la loi qui les unit n'est pas identique à la disjonction de lois qui relie ces dernières.

Cette distinction est d'importance, car le problème de la préemption de la causalité physique ne se pose que pour la version fonctionnaliste du physicalisme des particuliers, puisque lui seul refuse l'identification de la propriété mentale à une propriété physique de premier ordre, bien que ses deux formes soient non-réductionnistes au sens préalablement établi.

3. PROBLÈME DE L'EXCLUSION ET RECOURS À L'ÉMERGENTISME :

Ce problème ne se pose en fait que dans une approche réaliste du fonctionnalisme, selon laquelle la propriété mentale entendue comme une propriété

fonctionnelle possède une réalité objective, et du fait même une efficacité causale. Le propre du réalisme mental, tel notamment que Fodor s'en est fait l'impénitent défenseur contre l'approche de tendance instrumentaliste illustrée au contraire par Daniel Dennett (Dennett 1987), est en effet de considérer qu'une explication en termes mentaux est une explication de type causal, où la possession de certaines propriétés mentales est tenue pour causalement responsable de la possession d'autres propriétés mentales ou de manifestations comportementales. Dans la version fonctionnaliste du réalisme mental, l'efficacité causale de la propriété mentale devient donc nécessairement celle d'une propriété fonctionnelle. Et tout le problème de la préemption consiste en ce que cette efficacité causale, qui est essentielle à l'interprétation réaliste du fonctionnalisme, semble dérobée par celle des propriétés physiques dont la propriété mentale dépend en tant que propriété fonctionnelle.

Parmi les nombreux paramètres influant directement sur la position du problème, figure au premier chef la conception de la causalité mentale qui se trouve présupposée. Dans les faits, la difficulté se trouve principalement associée à une conception dite nomologique de la causalité²², héritée de Hume, et selon laquelle une relation de causalité n'est autre qu'une conjonction constante entre les instanciations de deux propriétés. Elle seule sera ici envisagée, sans exclure bien entendu en aucune façon par là qu'une analyse différente ou plus fine de la causalité permette de surmonter le problème de l'exclusion. Tout laisse au contraire à penser que c'est là que réside la clef de sa solution.

Dans la perspective qui est celle de la conception nomologique, l'efficacité causale n'appartient donc à proprement parler qu'à une instanciation de propriété et consiste dans le fait que l'instanciation d'un certain type de propriété est invariablement suivie de l'instanciation d'un autre type de propriété. Une propriété est dite par extension causalement efficace quand ses instanciations le sont au sens qui vient d'être précisé, c'est à dire donc quand ses instanciations sont ce en vertu de quoi survient invariablement l'instanciation d'une autre propriété. Il convient de distinguer de la notion de *propriété causalement efficace* ainsi entendue celle de *propriété causale*, qui désigne l'efficacité causale d'une propriété causalement efficace, c'est à dire encore des relations de causalité dans lesquelles elle peut entrer à titre de cause. On peut définir à son tour une *loi causale* comme une proposition énonçant une telle relation de conjonction constante entre deux propriétés, et une *explication causale* comme un argument subsumant une conjonction particulière d'événements sous une loi causale, ou une loi causale sous une loi causale plus générale. On dira par ailleurs qu'une propriété est *causalement pertinente* quand elle permet de formuler une loi causale et par là-même une explication causale. De cet ensemble de définitions semble découler le principe suivant, que l'on appellera le *principe de pertinence causale* : une propriété n'est causalement pertinente que si elle est causalement efficace. Un tel principe semble avoir l'évidence pour lui, dans la mesure où la fonction même d'une explication causale est manifestement de nous faire connaître la propriété en vertu de laquelle est instanciée une autre propriété.

Appliquant cette analyse de la causalité au cas des propriétés mentales, il apparaît que l'interprétation réaliste du fonctionnalisme considère la relation de causalité mentale comme une conjonction constante entre deux propriétés

²² Pour une formulation standard de cette idée à l'époque moderne, cf. par exemple Hempel (1987, §2.2), ou Carnap (1973, partie IV).

fonctionnelles, et à ce titre entre deux propriétés de propriétés physiques (cf. supra). De telle sorte que la loi qui figure au haut du schéma fodorien doit s'interpréter de la manière suivante : l'instanciation de la propriété M^1 est constamment conjointe avec l'instanciation de la propriété M^2 . Cette conception de la loi causale mentale appelle deux remarques.

La première est qu'une telle loi ne dit rien sur la nature physique de ce qui possède tant la propriété fonctionnelle mentale M^1 que la propriété fonctionnelle M^2 . En ce sens, c'est une loi causale abstraite, qui ne spécifie pas la nature dernière de ce qui possède la propriété causalement efficace M^1 ou de l'effet.

La seconde est que, dans la version causale du réalisme fonctionnel ainsi interprété, les propriétés fonctionnelles sont non seulement des propriétés causalement efficaces, mais aussi des propriétés causales, qui spécifient l'efficacité causale des propriétés physiques dont elles sont les propriétés. Ce qui ne va pas sans poser difficulté : une propriété causale peut-elle être causalement efficace ?

On la laissera toutefois ici de côté, dans la mesure où, bien qu'elle puisse ne pas être sans conséquences sur le problème de la préemption, elle n'est pas intrinsèquement liée avec lui²³. Celui-ci consiste en effet, ainsi qu'il apparaîtra dans les pages qui suivent, en ce qu'il semble à la réflexion illégitime de considérer les propriétés fonctionnelles comme des propriétés causalement efficaces en raison de leur statut de propriétés dépendantes de propriétés physiques, et non pas de leur caractère fonctionnel²⁴.

La raison en est qu'il paraît nécessaire de postuler que la propriété physique dont dépend la propriété fonctionnelle possède une efficacité causale identique à celle que possède la propriété fonctionnelle et qui a nécessairement la présence sur la sienne, la rendant du même coup au mieux redondante, et de ce fait inutile. En vertu même de sa dépendance par rapport à une propriété physique, une propriété fonctionnelle telle que M^1 ne peut être en effet instanciée sans qu'une propriété physique telle que P^1 ou P^2 ... le soit. Par ailleurs, le fait que M^1 soit causalement efficace par rapport à une autre propriété mentale M^2 signifie que chaque instanciation de M^1 est constamment suivie d'une instanciation de M^2 . Mais il en découle que chaque instanciation de P^1 ou P^2 ... est constamment suivie de l'instanciation de M^2 , et par voie de conséquence que P^1 ou P^2 ... sont elles-mêmes causalement efficaces par rapport à M^2 . Le même rapport de causalité qui existe entre M^1 et M^2 existe entre P^1 ou P^2 ... et M^2 . Qui plus est, le second est logiquement premier, puisque l'instanciation de P^1 ou P^2 est la condition nécessaire et suffisante de l'instanciation de M^1 .

Or si tel est le cas, l'efficacité causale de la propriété fonctionnelle est contestable. Et ce de deux façons principales. La première consiste à conclure que si l'efficacité causale d'une propriété telle que M^1 dépend nécessairement de celle d'une propriété P^x , et qu'elle est de plus identique à la sienne, alors c'est cette propriété P^x qui seule peut être causalement efficace, parce que M^1 n'a d'efficacité causale que pour autant que P^x en a. La seconde consiste à admettre qu'un tel argument n'établit que la redondance de l'efficacité causale de M^1 par rapport à celle de P^x , et qu'un autre argument est indispensable pour

²³ Elle est d'ailleurs largement ignorée; voir néanmoins à cet égard Block (1990).

²⁴ Il en existe de multiples versions, trop nombreuses pour être ici soumises à examen ; l'une des plus récentes est à ma connaissance celle de P. Humphrey (1997).

passer de la redondance à l'inexistence. Dans les deux cas, on peut dire que l'efficacité causale de la propriété physique préempte celle de la propriété mentale fonctionnelle, au point de la frapper au mieux d'inexistence et au pire d'inutilité. Si le raisonnement qui est proposé est correct, il me semble pour ma part préférable de retenir la première solution, qui revient à admettre qu'une relation de causalité est inexistante dès lors qu'elle peut être tenue pour inutile par rapport à une autre relation de causalité, et qui me paraît un critère d'existence tout à fait raisonnable pour une relation de causalité.

Il est évidemment essentiel à cet argument que le rapport de co-instanciation qui existe entre M^1 et P^x soit d'une autre nature que celui qui existe entre M^1 et M^2 . Sans quoi l'instanciation de M^1 ne jouerait au fond que le rôle d'une cause intermédiaire. Ce qui pourrait se produire de deux façons différentes : soit parce que les instanciations de M^1 et de P^x ne sont pas des instanciations simultanées, soit parce que - si tant est qu'il puisse y avoir quelque chose comme une causalité instantanée -, elles ne sont pas des co-instanciations logiques et correspondent à deux événements séparables. Au contraire, dans le cas de la propriété fonctionnelle, l'événement e que constitue son instanciation est aussi et du même coup l'instanciation d'une propriété physique²⁵. C'est une instanciation double.

Mais si une propriété mentale interprétée fonctionnellement se trouve privée de son efficacité causale, elle n'est plus, en vertu du principe de pertinence, causalement pertinente. Aussi les lois où elle figure ne sauraient-elles être des lois causales, et de la même façon, les explications qui font appel à ces dernières ne peuvent être tenues pour des explications causales. Il semble en effet qu'un énoncé tel que celui qui figure au sommet du schéma fodorien doive en réalité s'interpréter de la façon suivante : l'instanciation de certaines propriétés physiques, qui sont indéterminées sauf en ceci qu'elles possèdent telle propriété fonctionnelle M^1 , cause (est constamment conjoint à) l'instanciation de telles autres propriétés physiques, qui sont également indéterminées, sauf en ceci qu'elles possèdent telle autre propriété fonctionnelle M^2 . Or, d'une part, un tel énoncé ne dit pas que ce sont les propriétés fonctionnelles elles-mêmes qui sont en relation de causalité, puisque ce n'est pas en vertu d'avoir telle propriété fonctionnelle, mais certaines propriétés physiques, que l'entité x qui a la première cause la seconde. D'autre part, cet énoncé ne spécifie pas ce en vertu de quoi le rapport de causalité existe, c'est à dire la nature des propriétés causalement efficaces. Par là même, en vertu du principe de pertinence causale, il ne constitue pas une loi causale et les explications où il figure ne sauraient donc être tenues pour des explications causales.

Le problème de la préemption soulève de la sorte une double difficulté: 1) Existe-t-il un moyen de contrer l'argument qui établit l'exclusion de l'efficacité causale des propriétés mentales conçues comme propriétés fonctionnelles ? 2) Existe-t-il un moyen de maintenir une pertinence causale des propriétés mentales ainsi conçues même si l'on admet leur inefficacité causale, et par là même, un moyen de remettre en cause l'évidence apparente du principe de pertinence causale ?

C'est la première qu'il s'agit d'affronter ici, en examinant plus particulièrement si l'émergentisme est susceptible de fournir une forme spécifique de

²⁵ C'est en ce sens seulement que l'on peut parler d'identité d'instanciation dans le cas du matérialisme non-réductionniste de type fonctionnaliste.

matérialisme non-réductionniste qui permette ou d'éviter, ou de rendre aisément soluble, le problème de l'exclusion. Ce choix se justifie aisément par le fait que l'émergentisme s'est presque toujours présenté comme une forme de matérialisme récusant le réductionnisme. A l'époque moderne, il est en effet principalement associé, d'une part, au débat autour de l'autonomie de la biologie par rapport à la physique, qui naît dans les années 20 et s'étend jusqu'au début des années 60, et d'autre part, à celui sur l'autonomie de la psychologie par rapport à la neurobiologie au début des années 70. Or, dans les deux cas, il s'est avant tout défini à la fois comme une réaction contre les excès d'un matérialisme de type réductionniste, et comme une dénonciation des dangers d'une retombée dans le dualisme. En bref, comme un matérialisme non-réductionniste. Dès lors, la gravité des difficultés auxquelles semblent devoir faire face les versions fonctionnalistes et réalistes de ce dernier ne peut manquer de soulever l'interrogation suivante : une forme émergentiste du matérialisme non-réductionniste permet-elle d'échapper au problème de l'exclusion ? A la vérité, la question paraît même empreinte d'une certaine urgence, et l'on ne peut que s'étonner qu'elle n'ait jusqu'ici suscité qu'un intérêt fort limité²⁶, aussi bien de la part des partisans de l'émergentisme que de ceux du fonctionnalisme.

La réponse est simple dans son principe : ou l'élément responsable du problème de la préemption dans le matérialisme non-réductionniste de type fonctionnaliste est également présent dans l'émergentisme, à tout le moins dans certaines de ses formes, ou il en est absent. Dans le second cas, il reste alors à déterminer si un élément différent ne vient pas produire le même effet que lui. Il convient en conséquence de se pencher plus précisément sur les deux questions suivantes: a) Jusqu'à quel point, et en quoi, la notion d'émergence diffère-t-elle de la notion de fonctionnalisme ? b) Le cas échéant, cette différence est-elle essentielle, ou au contraire parfaitement secondaire, du point de vue du problème de la préemption ?

4. EXCLUSION ET ABSTRACTION :

41. La notion de propriété abstraite :

A cette fin, il est nécessaire de commencer par faire retour sur la notion de propriété fonctionnelle, et de tenter de déterminer avec une plus grande précision ce qui en elle est source du problème de l'exclusion.

En effet, selon les analyses proposées plus haut, la notion de propriété fonctionnelle possède, outre son caractère d'irréductibilité, quatre déterminations essentielles : a) c'est une propriété qui spécifie ce que fait l'entité physique qui la possède et non de quoi celle-ci est faite physiquement; b) c'est une propriété qui appartient à cette entité physique en vertu des propriétés qui la caractérisent en tant qu'entité physique, et qui par là est nécessairement dépendante quant à son existence et sa nature par rapport à elles; c) c'est aussi une propriété de ces propriétés physiques, et donc une propriété physique d'ordre supérieur; d) enfin, cette dépendance prend la forme spécifique d'une relation de survenance.

Or il ressort sans conteste des analyses précédentes que c'est manifestement la seconde de ces déterminations qui est aussi au cœur du problème de la préemption. C'est parce qu'il ne peut y avoir d'instanciation d'une propriété M sans

²⁶ Cf. Beckerman et al. (1992) ; et plus récemment Van de Vijver (1997).

co-instanciation d'une propriété P, que surgit le soupçon de l'inexistence ou de l'inutilité de l'efficacité causale de M. En d'autres termes, c'est l'affirmation "Pas de M sans P" qui porte la responsabilité de la difficulté. Je propose d'appeler les propriétés qui dépendent d'autres propriétés des propriétés abstraites, et les explications construites par leur moyen des explications abstraites²⁷. Aussi peut-on arrêter la définition suivante:

Une propriété abstraite est une propriété qui a pour condition nécessaire et suffisante d'instanciation la co-instanciation d'une autre propriété (ou de plusieurs autres).

Une telle terminologie se justifie du fait que, étant foncièrement dépendantes, de telles propriétés ne peuvent être isolées qu'en faisant abstraction de celles auxquelles elles sont liées, et par là même en introduisant un certain éloignement par rapport à la réalité, ainsi que le veut l'interprétation habituelle de l'opposition du concret et de l'abstrait. La propriété servant de condition nécessaire et suffisante sera désignée quant à elle du nom de propriété de base. Enfin, la relation de dépendance nécessaire et suffisante de la propriété abstraite par rapport à la propriété de base a pour converse la détermination de la propriété abstraite par la propriété de base.

A la vérité, malgré son rôle essentiel, la propriété d'abstraction ne saurait à elle seule porter la responsabilité du problème de l'exclusion. Il convient de lui adjoindre celle de l'irréductibilité. Car ce n'est en effet que parce que la propriété abstraite est irréductible à sa propriété de base qu'il peut y avoir un problème de l'exclusion: sans cela, l'efficacité causale de la première se réduirait tout simplement à celle de la seconde. On peut même aller plus loin encore, et soutenir que la notion de propriété abstraite n'a de sens véritable que pour autant qu'elle est irréductible. Dès lors en effet qu'elle est réductible, la dualité entre propriété abstraite et propriété de base en réalité disparaît. Pour autant, il serait en fait inadéquat d'intégrer la notion d'irréductibilité dans celle de propriété abstraite. Il paraît en effet indispensable de disposer d'un concept de propriété abstraite qui soit en lui-même neutre quant à la question de la réductibilité de cette dernière.

A partir de ces définitions, il est également possible de formuler l'hypothèse générale suivante, qui est en fait tout autant une formulation générale du problème de l'exclusion:

Toute propriété qui a pour condition nécessaire et suffisante d'instanciation la co-instanciation d'une autre propriété, et qui lui est irréductible, semble devoir être nécessairement privée de toute efficacité causale spécifique par cette autre propriété.

Mais dans quelle mesure ce caractère d'abstraction de la propriété fonctionnelle est-il indépendant des trois autres déterminations qui, au-delà de celle de l'irréductibilité, entrent dans la définition de son concept ? La question n'est pas sans importance. Car dans le cas où il ne l'est pas, le problème de l'exclusion ne peut surgir dans l'émergentisme sans que la notion de propriété émergente ne recouvre très largement celle de propriété fonctionnelle, sauf à provenir d'une source tout à fait différente.

²⁷ La notion de propriété abstraite est présente dans la littérature sur le problème de l'exclusion, en particulier chez Jackson et Pettit (cf. Pettit, 1990), mais elle demeure insuffisamment dégagée.

42. Hypothèse de l'indépendance de la notion de propriété abstraite :

Un certain nombre de raisons parlent clairement en faveur de l'hypothèse d'une telle indépendance de la détermination d'abstraction.

En premier lieu, en effet, une propriété peut manifestement dépendre nécessairement quant à son existence et sa nature d'une autre propriété, sans pour autant fournir une spécification fonctionnelle causale de ce à quoi elle se trouve de la sorte attribuée. La couleur peut ainsi être considérée comme une propriété nécessairement dépendante de celle de surface, sans qu'elle spécifie pour autant l'efficacité causale de ce qui est étendu.

De la même manière, une propriété n'a pas besoin d'appartenir à une autre propriété (d'être une propriété d'ordre supérieur) pour ne pouvoir être instanciée sans que cette propriété ne le soit. Il convient en effet de distinguer conceptuellement le cas où une propriété M ne peut appartenir à une entité x sans que cette entité possède une propriété P, mais où elle appartient néanmoins directement à x, et celui où une propriété M ne peut appartenir à une entité x sans que cette entité possède une propriété P, mais où elle appartient indirectement à x par le biais de son appartenance à P.

Le cas de la notion de survenance semble plus difficile à plaider, dans la mesure où celle-ci est habituellement associée à l'idée générale de dépendance ontologique. Mais elle n'en est au mieux qu'une forme spécifique, et différente de celle incarnée par le concept de propriété abstraite précédemment introduit. Sans pouvoir entrer ici dans le débat important qui entoure son analyse²⁸, il suffira pour faire apparaître ce point d'en rappeler la caractérisation classique offerte par Davidson dans « Événements mentaux », selon laquelle elle exprime l'idée « qu'il ne peut y avoir deux événements qui soient semblables sous tous leurs aspects physiques mais qui diffèrent sous un aspect mental quelconque, ou encore qu'un objet ne peut changer d'un point de vue mental sans changer d'un point de vue physique »²⁹. Encore que la relation ainsi définie puisse à mon sens s'interpréter comme une simple corrélation, il est manifestement présupposé par Davidson qu'elle est une relation de détermination, et donc de dépendance. Or en tant que telle, la caractérisation négative fournie dans cette citation implique en premier lieu qu'une propriété mentale est survenante par rapport à une propriété physique dans la mesure où cette propriété physique détermine cette seule propriété mentale, à l'exclusion de toute autre. Mais le point le plus important est qu'elle implique aussi en second lieu que plusieurs propriétés physiques différentes peuvent déterminer cette même propriété mentale. Ce qui, en termes de dépendance, revient à dire qu'une même propriété mentale peut dépendre de plusieurs propriétés physiques. Et c'est là l'élément propre à la détermination et la dépendance survenantes.

Or il n'est en aucune façon inscrit dans l'idée qu'une propriété a pour condition nécessaire et suffisante d'instanciation celle d'une autre propriété, que cette condition soit variable. Aussi la notion de propriété abstraite est-elle tout à fait compatible avec l'idée qu'elle dépende d'une seule propriété de base, et donc qu'elle ne soit pas survenante par rapport à elle.

Il importe toutefois de souligner que l'hypothèse que la notion de propriété abstraite est indépendante de celles de spécification fonctionnelle, de propriété

²⁸ Cf. notamment Kim (1993).

²⁹ Davidson (1970, tr. fr. modifiée, p. 286).

d'ordre supérieur et de survenance n'implique pas l'hypothèse réciproque. Et de fait, un certain nombre d'arguments parlent plutôt en faveur du contraire pour les deux premières.

Ainsi la fonction que possède une entité paraît-elle clairement dépendante de sa nature, puisqu'elle ne saurait détenir cette fonction que si il est possible de lui attribuer certaines propriétés substantielles déterminées. Pour reprendre un exemple fréquemment invoqué à titre d'illustration de la notion de propriété fonctionnelle, il n'est pas de carburateur en papier mâché. De la même façon, quand elles se trouvent conçues comme des propriétés spécifiant le rôle causal d'entités physiques, les propriétés mentales se trouvent nécessairement placées dans une relation de dépendance avec certaines propriétés physiques de ces entités³⁰. Une telle dépendance est toutefois parfaitement compatible avec l'admission d'un certain degré³¹ de variation dans ces dernières. Ainsi le schéma fodorien rapporté plus haut signifie-t-il qu'il ne saurait y avoir d'instanciation de la propriété M^1 sans qu'il y ait d'instanciation de la propriété P^1 , ou de la propriété P^2

Pareillement, une propriété ne saurait être d'ordre supérieur sans que son instanciation soit dépendante de celle d'une propriété dont elle est la propriété, et elle doit être instanciée dès lors que la propriété à laquelle elle appartient l'est.

Il n'en va pas de même pour la notion de survenance, qui semble plutôt être aussi indépendante de celle de propriété abstraite que celle-ci l'est d'elle-même. Compte non tenu du fait que, si elle est effectivement interprétable en termes de pure corrélation, la propriété survenante n'implique pas même les idées de dépendance et de détermination essentielles au concept de propriété abstraite, la définition davidsonienne, qui sert largement de référence au débat, n'exige en aucune façon qu'une propriété physique soit une condition nécessaire, et non pas simplement suffisante, de la propriété mentale qui survient par rapport à elle. En d'autres termes, elle est parfaitement compatible avec l'idée qu'une propriété mentale puisse être instanciée sans qu'aucune propriété physique le soit, même si elle admet que, dès lors que l'un des membres d'une disjonction de propriétés physiques déterminées l'est, elle l'est également.

Si tel est bien le cas, il semble que la notion de survenance ne saurait permettre de définir celle de physicalisme, ainsi qu'on le propose fréquemment aujourd'hui (cf. Braddon-Mitchell, 1996; Papineau, 1994), car cette dernière paraît au contraire inséparable de l'idée que des propriétés physiques soient des conditions nécessaires, et non pas seulement suffisantes, des propriétés mentales. Et elle saurait bien entendu d'autant moins le permettre si elle est en elle-même dépourvue de tout lien intrinsèque avec l'idée de détermination et de dépendance. J. Kim va par exemple dans le même sens, en faisant valoir que la définition davidsonienne revient simplement à poser une relation de détermination logique du type de celle qui est en jeu dans la notion mathématique de

³⁰ Considérée dans toute sa généralité, la notion de propriété fonctionnelle n'implique toutefois de dépendance que par rapport à une propriété substantielle, et non pas à une propriété substantielle *physique*. En ce sens, le fonctionnalisme n'est pas intrinsèquement matérialiste, ainsi que s'est employé à le souligner un de ses premiers avocats (cf. Hilary Putnam, *op.cit.*)

³¹ Contrairement à une idée reçue, il me semble patent que l'idée de fonctionnalisme n'implique aucunement que la variation des propriétés de base soit infinie, et par conséquent que l'on puisse « faire un esprit avec n'importe quoi » (cf. Roy, 1999).

fonction, et que cette dernière ne correspond pas à l'idée de détermination dont a besoin le matérialisme³².

Quoi qu'il en soit de ce dernier point, j'admettrai ici au minimum que la notion de dépendance impliquée par le concept de survenance ne correspond pas à celle qui est impliquée par le concept de propriété abstraite tel qu'il a été défini, et par voie de conséquence que ce dernier, qui est essentiel au caractère matérialiste du fonctionnalisme, n'est introduit que par l'idée de spécification fonctionnelle ou celle de propriété d'ordre supérieur.

43. Hypothèse de l'indépendance de la notion de propriété irréductible :

On a déjà admis l'indépendance de la notion de propriété abstraite par rapport à celle d'irréductibilité. Et il va de soi que cette indépendance est réciproque: deux propriétés peuvent être irréductibles l'une à l'autre sans être dans un rapport de propriété abstraite à propriété de base. Mais il est nécessaire d'examiner également si l'idée d'irréductibilité est indépendante des déterminations autres que celle d'abstraction qui entrent dans la définition de la propriété fonctionnelle. Car si elle ne l'est pas, une fois encore, le problème de l'exclusion ne peut surgir au sein de l'émergentisme que si la notion de propriété émergente recouvre en assez large part celle de propriété fonctionnelle.

Or il est patent que, de même qu'une propriété peut être abstraite sans être fonctionnelle, de second ordre ou survenante, elle peut être irréductible à une autre propriété sans qu'elle en spécifie la fonction, lui appartienne, ou survienne par rapport à elle. Réciproquement, une propriété fonctionnelle ou de second ordre n'est pas nécessairement irréductible. Le cas de la survenance est un peu plus complexe. On peut être tenté d'affirmer que le concept de physicalisme des types opposé par Fodor au réductionnisme *stricto sensu* (cf. *supra*) montre qu'il suffit de renoncer à l'idée que la disjonction des propriétés physiques par rapport auxquelles elle survient soit nomique, pour qu'une propriété mentale puisse être à la fois survenante et réductible. Mais il faudrait alors considérer que la propriété disjonctive est survenante par rapport aux propriétés qui la composent, ce qui n'est manifestement pas le cas. Si le concept de survenance implique donc celui d'irréductibilité, il importe néanmoins de souligner leur différence. En elle-même, la notion d'irréductibilité ne signifie rien d'autre qu'une impossibilité d'identification, tandis que la notion de survenance exprime une relation de détermination entre propriétés qui sert (entre autres) de fondement à l'affirmation d'une telle impossibilité d'identification.

De l'ensemble de ces considérations, il résulte qu'il convient en premier lieu de déterminer si la notion de propriété émergente contient, au même titre que celle de propriété fonctionnelle, le double élément d'abstraction et d'irréductibilité où le problème de l'exclusion puise sa source dans le fonctionnalisme. Or que celui d'irréductibilité en fasse partie, c'est là une thèse déjà admise en quelque sorte *ex hypothesi*, même si il demeure à l'établir, ainsi qu'à préciser la forme de cette irréductibilité. Aussi faut-il en fait s'attacher à déterminer avant tout si une propriété émergente est une propriété abstraite.

³² Cf. Kim (1993).

5. ÉMERGENCE, ABSTRACTION ET IRRÉDUCTIBILITÉ :

51. Définition de la notion d'émergence :

L'exigence de définir la notion de propriété émergente à laquelle on se trouve par là confronté se heurte à la grande variété des caractérisations qui en ont été proposées, et qui rendent nécessaire un travail d'analyse critique manifestement au-delà des limites de cet essai³³. Il est toutefois possible de dégager un concept provisoire d'émergence suffisamment général pour englober les principales formes prises par l'émergentisme depuis le XIX^e siècle, et cependant suffisamment rigoureux pour permettre la formulation d'une hypothèse précise quant à la question de l'inefficacité causale des propriétés émergentes.

En premier lieu, la propriété d'émergence est étroitement liée *de facto*, sinon *de jure*, à la relation de partie à tout, et se présente du fait même principalement comme une caractéristique du complexe. C'est ainsi que dès les premières définitions modernes, attribuées à John Stuart Mill³⁴ et George Henry Lewes³⁵, elle se trouve définie comme la propriété d'un tout. Car bien que ces derniers l'introduisent dans le cadre d'une réflexion sur la causalité, elle est toutefois directement motivée par le cas de la causalité complexe, c'est à dire celui où plusieurs relations de cause à effet se trouvent conjointement associées (« la composition des causes »). Ce que Mill et Lewes qualifient respectivement de « hétéropathique » et « émergent » est avant tout un effet complexe considéré dans son rapport avec les effets simples ou moins complexes dont il se compose.

En second lieu, le propre de la propriété émergente en tant que propriété d'un tout est d'être irréductible à la conjonction logique des propriétés de ses parties. Ainsi, reprenant l'analyse des rapports de causalité que Mill appelle dans sa *Logique* hétéropathiques, Lewes écrit-il dans *Problems of Life and Mind*: « Dans ce dernier cas, je propose d'appeler l'effet un émergent. L'émergent diffère de ses composants en ce qu'il est incommensurable avec eux, et il ne peut être *réduit* ni à leur somme ni à leur différence »³⁶. L'exemple le plus familier de propriété non émergente d'un tout, encore appelée *propriété résultante*, qui parcourt toute la littérature émergentiste est sans doute celui du poids du tas de sable, qui n'est rien d'autre que la conjonction de celui de chacun des grains entrant dans la composition du tas. Car même si cette conjonction prend alors la forme d'une propriété nouvelle (le poids spécifique du tas), celle-ci s'analyse exhaustivement comme la somme des poids des grains. Au contraire, une propriété émergente ne peut être analysée de la sorte. Pour autant, tous les cas de non émergence ne présentent pas la même homogénéité que dans cet exemple, où la différence entre le poids du tas et celui des grains est purement quantitative et n'implique aucune différence de type de propriété. Si les trois parties d'un tout possèdent respectivement les propriétés caractéristiques hétérogènes P¹, P² et P³, le tout qu'elles forment possède néanmoins nécessairement la propriété complexe P⁴, qui n'est rien d'autre que la conjonction (P¹ P² P³).

Une grande partie du débat sur l'émergentisme, notamment dans la querelle qui le mit aux prises avec le vitalisme et le mécanisme dans les années vingt, a

³³ Pour une intéressante tentative de classification des types de définition qui ont été proposées au cours de l'histoire moderne du concept, cf. Robert Klee (1984).

³⁴ Mill (1843).

³⁵ Lewes (1875).

³⁶ Lewes (1875, p. 412).

tourné autour de la manière dont il convient de définir positivement cette irréductibilité de la propriété du tout à la conjonction logique de ses parties. Deux de ses principaux protagonistes britanniques, Samuel Alexander et Lloyd Morgan, qui ont joué un rôle crucial dans le développement historique du concept d'émergence, ont proposé de la caractériser en termes d'imprédictibilité: une propriété d'un tout serait émergente dès lors qu'on ne pourrait la prédire sur la base de la connaissance exhaustive des propriétés de ses parties. Morgan écrit ainsi : « Ce qui ne peut être prédit, selon nous, est donc l'émergence de l'expression d'une certaine relation entre des événements préexistants. Quelqu'un doté de la seule connaissance de ces événements physico-chimiques, même la plus complète possible, ne permettrait de prédire le caractère des événements vivants qui émergent... »³⁷. Dans un article où il propose une brève analyse de l'évolution historique du concept d'émergence Stephan 1992³⁸, Achim Stephan affirme qu'il revient à un autre protagoniste de ce même débat, C.D. Broad, d'avoir substitué la notion de non-déductibilité à celle d'imprédictibilité. Ce dernier écrit en effet dans son ouvrage de 1925, *The Mind and its place in Nature*: « Formulée en termes abstraits, la théorie de l'émergence affirme que certains tous sont composés de constituants A, B, C, unis les uns aux autres par une relation R; que tous les tous composés de constituants de la même espèce que A, B et C unis entre eux par des relations de la même espèce que R possèdent certaines caractéristiques; que A, B et C sont capables de figurer dans d'autres types de complexes, où la relation qu'ils entretiennent l'un avec l'autre est d'une espèce différente; et que les caractéristiques du tout R(A,B,C) ne peuvent, même en théorie, se déduire de la plus complète connaissance des propriétés de A, B et C ou considérés isolément, ou figurant dans des tous qui ne sont pas de la forme R(A,B,C). La théorie mécanique rejette la dernière partie de cette affirmation »³⁹. Dans cette perspective, la propriété d'un tout est émergente dès lors qu'elle ne peut être logiquement déduite de la connaissance de celles de ses parties. C'est d'ailleurs en termes de déduction qu'elle se trouve encore caractérisée dans la célèbre définition que Carl Hempel et Paul Oppenheim ont fourni de la notion d'émergence en 1948, et qui sert de point de départ à une large partie des formes contemporaines de l'émergentisme, même si ces derniers la relativisent à un état de connaissance donné. Une telle définition est en effet manifestement plus satisfaisante, puisque dès lors qu'un lien régulier a été établi entre une propriété émergente et ses conditions d'émergence, la première peut être prédite sur la base du constat des secondes.

Certains se contentent cependant de rabattre l'opposition entre propriétés résultante et émergente sur celle entre propriétés homogène et hétérogène. La seconde opposition ne saurait toutefois être utilisée sans précaution pour expliquer la première. Si l'on entend en effet par homogène une propriété du tout qui, comme le poids dans un tas de sable, est partagée par chacune des parties, quoique dans une proportion différente, une propriété comme P⁴ doit être considérée comme hétérogène, puisque, bien qu'elle soit la conjonction logique de celles des parties du tout auquel elle appartient, elle ne satisfait pas cette exigence. C'est pourquoi Mario Bunge, qui est favorable à une caractérisation de l'émergence en termes d'hétérogénéité, prend soin de préciser qu'une pro-

³⁷ Morgan (1925, p. 5).

³⁸ Stephan (1992).

³⁹ Broad (1925, p 61).

priété d'un tout est résultante quand elle est partagée par au moins l'une des parties de ce tout, et émergente quand elle ne l'est par aucune d'entre elles⁴⁰. Il demeure qu'une propriété telle que P^4 ne peut *stricto sensu* être dite possédée à un degré différent par chacune de ses parties, mais seulement l'être partiellement par chacune d'elles. Aussi semble-t-il préférable, sans pour autant prétendre épuiser les mérites d'une telle approche, de retenir ici la définition en termes de non-déductibilité.

Ainsi entendue, la propriété émergente d'un tout se trouve, en troisième et dernier lieu, clairement placée dans une relation de dépendance avec les propriétés des parties de ce tout : et cette dépendance est telle que non seulement l'instanciation de la seconde entraîne l'instanciation de la première, mais de plus que la première ne saurait être instanciée sans que les secondes le soient. Comme l'exprime fort bien la définition de C. D. Broad, la caractéristique du tout $R(A,B,C)$ ne peut apparaître que si les éléments A, B et C sont présents et liés par la relation R. En d'autres termes, les parties du tout sont des conditions nécessaires et suffisantes de l'émergence de la propriété du tout en question.

L'ensemble de ces déterminations peuvent être regroupées au sein de la définition formelle suivante, inspirée de Achim Stephan :

La propriété d'un tout (système) est dite absolument émergente si

1) Loi L^1 de transition : Il existe une loi telle que chaque fois qu'une série de conditions $B^1... B^n$ sont satisfaites de façon simultanée, un ensemble d'éléments $C^1... C^n$ entrent dans une relation R et constituent de la sorte une microstructure $Ms = [C^1... C^n ; R]$;

2) Loi L^2 de Propriété : Chaque fois que la microstructure $[C^1... C^n]$ est réalisée, une propriété P est instanciée, et la propriété P ne peut être instanciée sans que cette microstructure soit réalisée ;

3) L'existence de la propriété P ne peut être déduite d'aucune loi portant sur les éléments de la microstructure considérés isolément ou au sein d'autres microstructures Ms', Ms''

Cette définition peut aisément être relativisée : il suffit de modifier la condition 3 et de préciser que P ne peut être déduite d'aucune loi de telle ou telle théorie T.

52. Émergence et abstraction :

Or il est clair que la notion de propriété émergente ainsi définie satisfait pleinement, par la condition 2, la définition du concept de propriété abstraite: les conditions d'émergence représentent autant de propriétés qui sont co-instanciées avec la propriété dite émergente, et dont l'instanciation est tout à la fois une condition nécessaire et suffisante de la sienne.

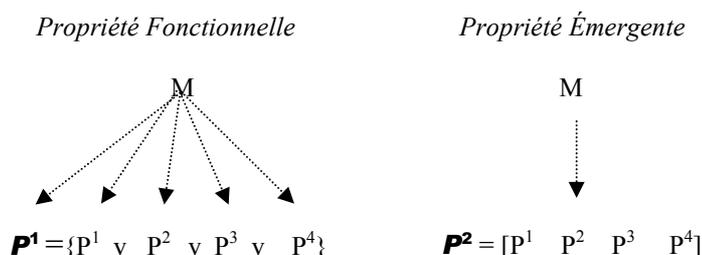
Et il n'est pas moins confirmé qu'elle contient l'élément d'irréductibilité requis par le problème de l'exclusion. Jusqu'à quel point, cependant, la notion d'irréductibilité mise en jeu est-elle semblable à celle qui se trouve associée au concept de propriété fonctionnelle?

Or il convient ici de nettement distinguer non seulement, une fois encore, la question de la nature de l'irréductibilité et celle de son fondement, mais aussi celle de sa base, au sens de ce à quoi la propriété mentale est dite irréductible, et qui est également ce dont elle dépend, c'est à dire la base de son abstraction.

⁴⁰ Mahner et Bunge (1997).

Quant au premier point, les deux notions ne présentent aucune différence: il s'agit dans les deux cas d'une impossibilité d'identification. Elles diffèrent néanmoins sur les deux autres.

Une propriété fonctionnelle est en effet irréductible, d'une part, à chacune d'un ensemble de propriétés simples, et d'autre part, à la propriété complexe de type disjonctif que forment ces propriétés simples. Au contraire, dans le cas de l'émergentisme, la base est constituée par une propriété complexe et conjonctive, puisque la relation d'émergence est inscrite dans le cadre d'un rapport entre tout et parties. Cette différence, qui est tout autant une différence de type de propriété abstraite, peut se résumer au moyen du schéma suivant, où chaque flèche en pointillés représente une relation d'irréductibilité :



Différence de bases de réduction

En second lieu, l'irréductibilité se trouve justifiée de manière différente dans chacun des deux cas. Celle de la propriété fonctionnelle résulte tout à la fois de la pluralité des propriétés de base simples (argument de la réalisabilité multiple) et du caractère anomique, tant de la disjonction qu'elles forment, que de la corrélation entre cette disjonction et la propriété fonctionnelle. Rien de tel pour la propriété émergente, puisque ni la pluralité des propriétés de base simples ni la nomicité de la propriété complexe conjonctive qu'elles forment ne sont en cause; seule l'est la possibilité d'établir un rapport de déduction. La propriété émergente est irréductible à ses conditions d'instanciation simplement parce qu'elle n'est pas déductible à partir d'elles.

Il ressort donc de l'ensemble de ces analyses que propriété fonctionnelle et propriété émergente constituent en réalité deux espèces de propriété abstraite et irréductible, quoique deux espèces différentes. Or rien dans ce qui fait la particularité de l'irréductibilité et de l'abstraction de la propriété émergente (nature de la base de son abstraction et de son irréductibilité, ainsi que fondement de son irréductibilité) ne paraît manifestement en mesure de prévenir le surgissement du problème de l'exclusion, confirmant ainsi l'hypothèse générale proposée plus haut, selon laquelle toute propriété abstraite et irréductible est apparemment privée d'efficacité causale par ses conditions d'instanciation.

6. CONFRONTATION DES NOTIONS DE PROPRIÉTÉS ÉMERGENTE ET FONCTIONNELLE :

Il reste toutefois possible en principe que les conséquences épiphénoménales de ce statut de propriété abstraite et irréductible se trouvent bloquées par la présence, au sein de la notion de propriété émergente, d'autres déterminations essentielles absentes de l'essence de la notion de propriété fonctionnelle, ou du moins qu'elles soient aisément contournables grâce à celles-ci. Aussi est-il

nécessaire de cerner avec plus de précision ce qui fait, au-delà des différences qu'elles présentent au niveau des déterminations d'abstraction et d'irréductibilité, l'essence spécifique de chacun des deux types de propriétés, puis de déterminer jusqu'à quel point les différences spécifiques additionnelles ainsi mises à jour sont susceptibles de mettre en cause la solution sur laquelle débouchent les analyses précédentes.

61. Hypothèse sur leur spécificité respective :

Il convient à cet effet de commencer par recenser les principales caractéristiques de chacune des deux propriétés dégagées jusqu'ici et de les mettre en regard les unes avec les autres. Ce qui peut être réalisé au moyen du tableau suivant, où les signes centraux d'identité et de différence soulignent les points de divergence et de convergence entre déterminations, tandis que le terme "Non" indique une absence de détermination correspondante:

	A: PROPRIETE FONCTIONNELLE		B: PROPRIETE EMERGENTE
1	Abstraction	=	Abstraction
2	Irréductibilité :	=	Irréductibilité :
a	- nature : non-identifiabilité	=	- nature : non- identifiabilité
b	- fondement : variabilité et anomicité	#	-fondement : non-déductibilité
3	Base de l'abstraction et de l'irréductibilité :	#	Base de l'abstraction et de l'irréductibilité :
	v propriétés		de propriétés
4	<i>Non</i>	#	Propriété d'un tout
5	Spécification fonctionnelle	#	<i>Non</i>
6	Survenance	#	<i>Non</i>
7	Ordre supérieur	#	<i>Non</i>

Il faut ensuite déterminer dans quelle mesure chacune de ces déterminations appartient à l'essence des propriétés qu'elles spécifient respectivement, puis dégager sur cette base les différences qui séparent ces dernières l'une de l'autre. Il est à remarquer à cet égard que les "non" ne traduisent que de simples défauts de mention, et que la réalité effective de l'absence qu'ils expriment reste par conséquent à établir. Par ailleurs, si une différence de déterminations essentielles s'oppose à une communauté de déterminations essentielles, elle n'implique en aucune façon cependant que les déterminations qu'elle met en jeu ne puissent appartenir à titre non essentiel à la propriété de l'essence de laquelle elles sont respectivement exclues. Une différence essentielle n'est, en d'autres termes, pas nécessairement exclusive. Elle ne l'est qu'à la condition qu'au moins une des déterminations concernées ne puisse appartenir à la propriété de l'essence de laquelle elle est exclue, sans que cette propriété soit dénaturée. Ce qui peut advenir parce que celle-ci cesse ou d'être ce qu'elle est, ou d'être purement ce qu'elle est.

Or l'ensemble des analyses précédentes semble fournir en fait déjà la plupart des éléments de réponse nécessaires. En effet:

I) Il vient d'être établi qu'il est essentiel à la propriété émergente de posséder les déterminations (B) 1, 2 a-b, et 3.

II) Il a également été établi que les déterminations (A) 1, 2 a-b et 3 appartiennent de manière essentielle à la notion de propriété abstraite et irréductible liée à celle de propriété fonctionnelle.

III) Il est de plus manifeste qu'il n'est en aucune façon nécessaire à une telle forme de propriété abstraite et irréductible de type fonctionnel d'appartenir à un tout (4), sans impliquer pour autant par là qu'elle ne peut y appartenir.

IV) Par définition même, une propriété fonctionnelle doit spécifier une fonction (5), exigence à laquelle n'est par contre pas soumise une propriété émergente en tant que telle.

Il apparaît déjà ainsi que les différences 2 b, 3, 4 et 5 entre A et B sont essentielles, et il ne reste en conséquence qu'à examiner les deux dernières mentionnées dans le tableau.

Or s'il a été établi que les concepts de propriété abstraite et de propriété survenante étaient indépendants l'un de l'autre, il n'en demeure pas moins que les déterminations du second sont toutes présentes dans le type de propriété abstraite et irréductible lié à l'idée de propriété fonctionnelle. Aussi cette sixième différence est-elle également essentielle.

Il a enfin été admis qu'une propriété fonctionnelle était une propriété d'ordre supérieur, quoiqu'une propriété abstraite n'en soit pas nécessairement une. Ceci laisse assurément entière la question de savoir si elle l'est elle-même de façon nécessaire, et si elle diffère en ceci de la propriété émergente. Pourtant, dès lors que l'on accepte le principe qu'une propriété abstraite puisse ne pas être d'ordre supérieur, il n'y a aucune raison de refuser cette possibilité de principe à ses espèces particulières qu'en sont les propriétés fonctionnelle et émergente. Et ce, même si il semble particulièrement naturel d'affirmer que, si une entité *x* possède une fonction en vertu de certaines propriétés substantielles, cette fonction lui appartient en tant qu'elle a ces propriétés, et donc par le biais de ces propriétés. La dernière différence du tableau sera en conséquence tenue ici pour inessentielle. En tout état de cause, on ne voit pas en quoi le fait d'être essentielle permettrait à la propriété émergente d'échapper au problème de l'exclusion.

A l'exception de cette dernière, il semble donc au total que l'ensemble des différences mentionnées au tableau sont de nature essentielle, et que celui-ci offre donc d'emblée une image fidèle de la spécificité des deux propriétés en jeu.

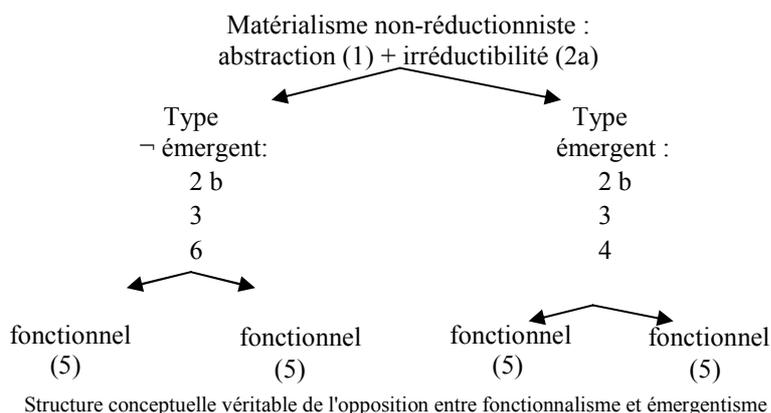
62. Objection à l'hypothèse :

Cette conclusion s'avère pourtant à la réflexion contestable. Elle repose en effet sur la supposition qu'il existe un lien intrinsèque entre la détermination de fonctionnalité et les autres déterminations spécifiques auxquelles elle est liée (à savoir 2b, 3 et 6) Mais une telle supposition est tout à fait contestable. En effet, si une propriété fonctionnelle doit bien par définition même avoir la détermination de fonctionnalité, et d'autre part celle d'être abstraite et non identifiable avec ce dont elle dépend, rien ne semble devoir faire obstacle cependant à ce qu'elle soit une propriété abstraite et irréductible d'un autre type, par exemple celui représenté par la notion de propriété émergente. Et du même coup qu'une propriété émergente puisse être fonctionnelle, en ce qu'elle ait la détermination de fonctionnalité. Il convient par conséquent de modifier le

point IV ci-dessus, et d'affirmer d'une part que, si il est essentiel à la notion de propriété fonctionnelle d'avoir la détermination de fonctionnalité, il n'est pas essentiel au type de propriété abstraite désigné sous ce nom d'être fonctionnelle; et d'autre part, qu'il n'est pas essentiel à une propriété émergente de ne pas être fonctionnelle. Or une telle rectification a des conséquences de première importance sur l'analyse de ce qui fait la différence spécifique entre une propriété fonctionnelle et une propriété émergente.

Il apparaît en effet que les deux notions qui viennent d'être comparées ne sont pas en réalité de même niveau. Leur différence recouvre certes bien une différence entre deux types de propriétés abstraites et irréductibles. Mais ce qui se trouve ici désigné du nom de propriété fonctionnelle n'est qu'une sous-espèce de la première, et il existe aussi une sous-espèce de la seconde qui mérite non moins le titre de propriété fonctionnelle. On ne peut donc opposer en tant que telle la notion de propriété fonctionnelle et celle de propriété émergente, puisque aussi bien une propriété émergente peut être fonctionnelle, et une propriété désignée ici comme fonctionnelle ne pas l'être. On ne peut donc se satisfaire non plus de la manière dont la différence entre le matérialisme non réductionniste de type fonctionnaliste est habituellement conçue, puisqu'elle revient à présenter comme une différence d'espèce ce qui est en réalité une différence entre une espèce et une sous-espèce d'une autre espèce.

Pour mieux faire apparaître la véritable géographie conceptuelle en jeu dans l'opposition entre les matérialismes non-réductionnistes fonctionnaliste et émergentiste, il convient donc de substituer au tableau précédent le schéma arborescent suivant:



Il est à noter que chacun des quatre types ainsi distingués peut en outre être ou non d'ordre supérieur.

Or du même coup, il apparaît aussi qu'il n'y a rien de plus dans l'essence spécifique de la notion d'émergence que ce qui avait déjà été examiné. Par voie de conséquence, la possibilité de principe que d'autres déterminations spécifiques lui permettent de bloquer ou de contourner le problème de l'exclusion se trouve définitivement écartée, ainsi que toute cette analyse avait pour souci de le déterminer.

L'opposition entre propriété fonctionnelle et propriété émergente recouvre donc au total une différence entre deux types spécifiques de propriété abstraite et irréductible, mais qui chacun paraît égal devant le problème de l'exclusion.

Références

- Alexander, S. (1920). *Space, Time and Deity*. London: MacMillan.
- Beckermann, A., Flhor, H. et Kim, J. (1992). *Emergence or reduction: Essays on the prospects of Nonreductive Physicalism*. Berlin: Walter de Gruyter.
- Borst, C.V. (1970). *The Mind-Brain Identity Theory*. London: MacMillan.
- Block, N. (1980). *Readings in the Philosophy of Psychology*. Vol 1, p 207-215. Cambridge: Cambridge University Press.
- (1990). Can the mind change the world? In George Boolos (ed.): *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge: Cambridge University Press.
- Braddon-Mitchell, D. et Jackson, F. (1996). *Philosophy of Mind and Cognition*. Oxford: Basil Blackwell.
- Broad, C.D. (1925). *The Mind and its place in Nature*. London: Routledge & Kegan Paul.
- Carnap, Rudolf. (1928). *Der Logische Aufbau der Welt*. Berlin.
- (1932-33). Psychologie in Physikalischer Sprache. *Erkenntnis*.
- (1973). *Les problèmes de la philosophie de la physique*. Traduit de l'anglais par J.M. Luccioni et A. Soulez. Paris: Armand Colin.
- Davidson, D. (1970). *Mental Events, Actions and Events* (1980). Oxford: Oxford University Press.
- Davidson, D. (1973). *The Material Mind, Actions and Events* (1980). Oxford: Oxford University Press.
- Davidson, D. (1974). *Psychology as Philosophy, Actions and Events* (1980). Oxford: Oxford University Press.
- Davidson, D. (1980). *Actions and Events*. Oxford: Oxford University Press.
- Fodor, J.A. (1968). *Psychological explanation: an introduction to the philosophy of psychology*. New York: Random House.
- Fodor, J.A. (1975). *The Language of Thought*. Cambridge: Cambridge University Press.
- Hempel, C. et Oppenheim, P. (1948/1965). Studies in the Logic of Explanation. In C. Hempel (Ed.): *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: The Free Press.
- Hempel, C. (1987). Aspects of scientific explanation. In C. Hempel (Ed.): *Aspects of Scientific Explanation and other Essays*. New York: The Free Press.
- Humphrey, P. (1997). How properties emerge. *Philosophy of Science*, Vol 64, N°1.
- Jacob, P. (1980). *De Vienne à Cambridge*. Paris: Gallimard.
- (1991). Are mental properties causally efficacious? *Grazer Philosophische Studien*, 39, pp 57-73.
- Kim, J. (1993). *Supervenience and mind: selected philosophical essays*. Cambridge studies in philosophy (pp. xviii, 377). New York: Cambridge University Press.
- Klee, R. (1984). Micro-determinism and concepts of emergence. *Philosophy of Science*, 51, pp. 44 -63.
- Koenig, O. et Kosslyn, S. (1995). *Wet Mind. The New Cognitive Neuroscience*. New York: The Free Press.
- Lewes, G. H. (1875). *Problems of Life and Mind*. London: Kegan Paul.
- Lewis, D. (1966/1983). An argument for the Identity Theory. In: *Philosophical Papers*. Vol. 1. Cambridge: Cambridge University Press.
- Lewis, D. (1972). Psychophysical and Theoretical Identifications. In: *Philosophical Papers*. Vol. 1. Cambridge: Cambridge University Press.
- Lewis, D. (1983). *Philosophical Papers*. Vol. 1. Oxford: Oxford University Press.

- Macdonald, C. et Macdonald, G. (1995). *Philosophy of Psychology: Debates on Psychological Explanation*. Vol. 1. Oxford: Basil Blackwell.
- Mahner, M. et Bunge, M. (1997). *Foundations of Biophilosophy*. Berlin: Springer.
- Mill, J. S. (1843). *System of Logic*. London: Longmans.
- Nagel, E. 1979 (1960). *The structure of Science*. Hackett Publishing Company.
- Papineau D. (1994). *Philosophical Naturalism*. Basil Blackwell.
- Pettit, P. et Jackson, F. (1990). Program explanation: a general perspective. *Analysis*, 50, 107-117.
- Pepper, S. (1926). Emergence. *Journal of Philosophy*, pp 241-245, Vol 23, N° 9.
- Place, U.T. (1956/1970). Is consciousness a brain process? In C.V. Borst (Ed.): *The mind brain identity theory*. London: Macmillan.
- Popper, K. et Eccles, J. (1977). *The Self and its Brain*. Routledge. London.
- Putnam, H. (1967). The nature of mental states. In: *Collected Papers*. Vol. 1. Cambridge: Cambridge University Press. (1975). (Tr. fr. de J.M. Roy sous le titre "La nature des états mentaux" dans *La théorie computationnelle de l'esprit*, Les Études Philosophiques. 1993).
- Quine, W. (1960). *Word and Object*. Cambridge (MA): MIT Press.
- Roy, J-M. (1999). L'explication de style causal. Communication au département de Philosophie de l'Université de Toulouse, à paraître.
- Rorty, R. (1970). Mind-Body identity, privacy and categories. In C.V. Borst (Ed.): *The mind brain identity theory*. London: Macmillan.
- Stephan, A. (1992). The Historical Facets of Emergence. in Beckerman 1992.
- Van de Vijver, G. (1997). Emergence and Explanation. *Intellectica*, N°25.