

## Matérialisme et réduction de l'esprit

Max Kistler

► **To cite this version:**

Max Kistler. Matérialisme et réduction de l'esprit. Jean Dubessy, Guillaume Lecointre, Marc Silberstein. Les matérialismes (et leurs détracteurs), Syllepse, pp.309-339, 2004. <ijn\_00000551>

**HAL Id: ijn\_00000551**

**[https://jeannicod.ccsd.cnrs.fr/ijn\\_00000551](https://jeannicod.ccsd.cnrs.fr/ijn_00000551)**

Submitted on 29 Nov 2004

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Matérialisme et réduction de l'esprit

Max KISTLER

### 1 – Arguments contre l'identité esprit-cerveau et contre la réductibilité de l'esprit

Une partie importante du débat sur la place de l'esprit dans la nature s'articule autour des arguments avancés par Hilary Putnam, Jerry Fodor et Donald Davidson dans les années 1960 et 1970. Ces arguments visent à établir premièrement que l'esprit *n'est pas identique* au cerveau et que les états mentaux ne sont pas identiques aux états du cerveau, et que deuxièmement, l'esprit et les états mentaux sont *irréductibles*, c'est-à-dire ne peuvent pas faire l'objet d'une réduction aux propriétés du cerveau.

Selon une hypothèse avancée par Hilary Putnam et qui a eu une influence considérable, les états mentaux sont des états fonctionnels : leur identité est déterminée par le rôle causal qu'ils jouent dans un organisme. À titre d'exemple, « l'état d'éprouver de la douleur [...] est un état fonctionnel d'un organisme tout entier » (Putnam, 1967, p. 154, trad. p. 276). Sans entrer dans les détails techniques, l'idée principale de la conception « fonctionnaliste » de l'esprit suggérée et développée par Putnam dans les années 1960 et 1970 est celle-ci<sup>1</sup> : ce n'est pas l'incarnation matérielle dans un cerveau avec ses propriétés physico-chimiques qui détermine la nature d'un état mental, mais plutôt les causes et effets que cet état mental a sur les autres états mentaux, les sensations (lors-

1. Pour plus de détails, voir Jacob (1992) et Pacherie (1993) qui offrent une présentation accessible des idées principales du fonctionnalisme.

qu'il s'agit d'un état perceptif), et le comportement de l'organisme. Cette thèse a pour conséquence que la matière dans laquelle l'esprit est incarné ne contribue absolument pas à déterminer la nature des états mentaux ; il est indifférent, dit Putnam, qu'il s'agisse de « chose-animée [*soul-stuff*], de matière ou de fromage suisse » (Putnam, 1975, p. 143), dans la mesure où ce substrat permet aux états mentaux d'avoir les relations causales et fonctionnelles qui déterminent seules leur identité.

Putnam tire deux conséquences importantes de son hypothèse : la première est que, si la nature des états mentaux est fonctionnelle et indépendante du substrat dans lequel ils sont « implémentés »<sup>2</sup>, alors « nos états mentaux, par exemple *penser aux vacances de l'été prochain*, ne peuvent être *identiques* à aucun état physique ou chimique » (Putnam, 1975, p. 135, italiques de Putnam). La raison est que « la réalisation de l'état mental est en un sens assez accidentelle » (Putnam, 1975, p. 136) : dans la mesure où le cerveau incarne un programme dont les états mentaux sont les états computationnels, nous savons qu'il est « physiquement possible, quoique non nécessairement pratiquement faisable, de produire quelque chose avec ce même programme mais une constitution physique et chimique assez différente » (Putnam, 1975, p. 135-136). L'état mental ne peut être identique à aucun état cérébral parce que, du moins en principe sinon dans des sujets réels, le même état mental peut être « réalisé » par différents états cérébraux. Or ces états cérébraux ne peuvent pas être à la fois différents entre eux et tous identiques à l'état mental qu'ils réalisent. C'est la fameuse thèse de la « réalisabilité multiple » (ou « multiréalisabilité ») qui prétend démontrer que, en vertu de sa nature fonctionnelle, un état mental ne peut être identique à aucun état cérébral particulier qui le réalise.

La seconde conséquence que Putnam tire de son hypothèse « fonctionnaliste » sur la nature des états mentaux, c'est qu'il est impossible de *réduire* les états mentaux à des états physico-chimiques du cerveau. Selon le modèle classique de la réduction dans la tradition de l'empirisme logique en philosophie des sciences, élaboré notamment par Nagel (1961), une réduction est une relation entre deux théories qui est fondée sur des principes de liaison (aussi appelés « principes pont » – *bridge principles* – ou « lois-pont » – *bridge laws*) exprimant l'équivalence exten-

2. Au lieu de dire, dans le jargon de la programmation informatique, que les états mentaux sont « implémentés » dans la matière cérébrale, on dit aussi qu'ils sont « réalisés » par la matière cérébrale. Nous reviendrons sur l'interprétation de cette forme de « réalisation ».

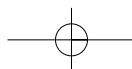


sionnelle des prédicats de la théorie réduite à des prédicats de la théorie réductrice : la réduction d'une théorie de haut niveau, par exemple biologique, à une théorie de plus bas niveau, par exemple chimique, passe par la déduction des lois de la théorie réduite (ici, la biologie), à partir des lois de la théorie réductrice (ici, la chimie), et des principes pont qui mettent en rapport les vocabulaires des deux théories<sup>3</sup>. Selon Putnam, la raison pour laquelle les propriétés mentales sont en principe irréductibles à des propriétés de plus bas niveau, par exemple neurophysiologiques, est qu'une telle réduction ne pourrait pas être explicative, dans la mesure où elle serait surchargée de détails sans pertinence ; par conséquent, aucune déduction à partir de la description physique du cerveau ne pourrait dégager les régularités du niveau psychologique qui sont constitutives de l'identité des états mentaux. De l'impossibilité analogue à propos des propriétés et lois économiques, Putnam dit que « la dérivation des lois de l'économie à partir des seules lois de la physique est *en principe impossible* » (Putnam, 1975, p. 139). Cependant, la différence entre la réduction et l'identité n'est pas toujours clairement perçue, dans la mesure où, selon l'une des principales conceptions de la réduction entre théories de différents niveaux<sup>4</sup>, la seule manière d'accomplir une réduction passe par *l'identification* des propriétés auxquelles font référence les prédicats de la théorie réduite à des propriétés auxquelles fait référence la théorie réductrice. Or, à condition de faire abstraction de cette conception particulière de la nature de la réduction, il apparaît que les thèses de la non-identité et de l'irréductibilité sont indépendantes. Je défendrai en effet l'idée selon laquelle les propriétés mentales sont *réductibles* à des propriétés appartenant à des niveaux inférieurs au niveau psychologique, à savoir à des propriétés neurophysiologiques, *sans être identiques* avec elles.

L'argument anti-réductionniste offert par Fodor ouvre la voie pour une telle conception, grâce à une différence par rapport à l'argument de Putnam qui passe souvent inaperçue. Selon Fodor, les prédicats désignant les espèces naturelles des sciences spéciales, c'est-à-dire des sciences étudiant les propriétés non-physiques, sont les prédicats « dont les termes sont les variables liées » (Fodor, 1974, p. 132, trad. p. 424) des lois de

3. Sur le concept de loi, cf. Armstrong (1993), Kistler (1999), et Barberousse, Kistler et Ludwig (2000, chap. 4).

4. Elle a notamment été défendue par Causey (1977).



ces sciences<sup>5</sup>. Ces prédicats s'appliquent directement aux objets; en ce sens, ce sont des prédicats «de premier ordre»; cela les distingue des prédicats fonctionnels parmi lesquels Putnam compte les prédicats psychologiques. Comme l'a montré Kim (1998), la forme logique de tels prédicats fonctionnels est de *second ordre*: être dans un état fonctionnel F dont l'identité est déterminée par le fait qu'il est causé par des états  $F_1$  et qu'il cause des états  $F_2$ , est équivalent au fait d'avoir *une propriété ou autre* qui est causé par  $F_1$  et qui cause  $F_2$ . Pour faire apparaître plus explicitement le fait que le prédicat fonctionnel F est de second ordre, on peut dire: selon une conception fonctionnaliste des états mentaux, le fait que x est dans l'état F est équivalent au fait qu'*il existe* une propriété (de premier ordre) P de x, telle que (l'instance de<sup>6</sup>) P est causé par (l'une des propriétés réalisant)  $F_1$  et que P cause (l'une des propriétés réalisant)  $F_2$ . La conception fonctionnaliste est construite «sur mesure» pour tenir compte de la possibilité de la réalisabilité multiple: une propriété conçue selon son rôle fonctionnel peut être «réalisée» par différentes, et en fait par un nombre indéterminé de propriétés de premier ordre, d'une manière analogue de celle dont un programme informatique peut être implémenté dans un nombre indéterminé de machines de structures physiques différentes.

En revanche, la conception – explicitement adoptée par Fodor – selon laquelle l'identité des propriétés est déterminée par les lois de la nature dans lesquelles elles participent, prépare le terrain pour une toute autre conception de la réalisabilité multiple et de la réduction. Si les propriétés, à tous les niveaux, se distinguent par les relations nomiques qu'elles entretiennent les unes avec les autres, nous sommes fondés de concevoir toutes ces propriétés comme des propriétés de *premier ordre*. Prenons la propriété d'un organisme de se représenter l'endroit où se trouve une cerise qu'il a l'intention de manger. Admettons que dans une situation donnée, cette représentation donne lieu, sur l'arrière-plan de l'intention de la manger et d'autres états mentaux, en particulier de désirs et moti-

5. Fodor (1974, p. 132 et p. 141, trad. p. 424 et p. 434). Cette conception a été généralisée à toutes les propriétés naturelles par Shoemaker (1980) et Kistler (2002).

6. Dans la mesure où une propriété est conçue comme un *universel*, elle ne peut ni causer ni être causée; ce sont les *instances* de la propriété, dans des objets ou événements localisés dans l'espace et dans le temps, qui sont causées et qui causent à leur tour des effets. La propriété universelle d'être une charge électrique en mouvement n'est ni une cause ni un effet. Mais l'instance de cette propriété dans l'électron particulier e, à un certain moment et à un certain endroit, est causée, disons par la proximité d'une charge positive qui l'attire, et elle cause à son tour un champ magnétique autour d'elle.



ventions, à l'action consistant à saisir le fruit et le mettre à la bouche. Le lien entre l'état mental – qui contient la représentation et l'intention – et l'action est nomique<sup>7</sup>. En concevant la propriété mentale comme terme d'une loi causale, on la conçoit comme une propriété de premier ordre, alors que la conception fonctionnaliste putnamienne revient à la concevoir comme une propriété de second ordre<sup>8</sup>. Cette différence dans la conception de la nature logique de la propriété est indépendante du jugement selon lequel elle appartient, comme toutes les propriétés psychologiques, au *niveau* des propriétés appartenant aux organismes : la propriété  $P_1$  appartient à un niveau plus élevé que la propriété  $P_2$  si le possesseur de  $P_2$  est une partie propre du possesseur de  $P_1$ . En ce sens, les propriétés psychologiques qui appartiennent à l'organisme entier se situent sur un niveau plus élevé que les propriétés neuronales qui appartiennent à des réseaux neuronaux, à des neurones individuels et les propriétés chimiques qui appartiennent aux molécules composant ces neurones, car les réseaux neuronaux, les neurones et les molécules sont des *parties* de l'organisme<sup>9</sup>.

La conception des propriétés mentales comme propriétés de premier ordre mais de plus haut niveau que les propriétés neuronales, ouvre la perspective d'une nouvelle interprétation de la réalisabilité multiple. Selon celle-ci, la réalisabilité multiple caractérise un rapport entre différentes propriétés réductrices de bas *niveau* capables de réduire *une* propriété de haut *niveau*, au lieu de caractériser, comme dans la conception de Putnam (et de Kim), un rapport entre différentes propriétés de premier *ordre* – de haut niveau – capables de réaliser *une* description de second *ordre* et de *même (haut) niveau*. La propriété réduite est une propriété macroscopique de premier ordre, par exemple une propriété biologique

7. Cela est parfaitement compatible avec le fait que la régularité qui l'accompagne a des exceptions, ce qui est vrai de la plupart des lois de la nature, même physiques. Sur les exceptions aux lois, cf. Hempel (1988), Kistler (1999 et 2003). La thèse selon laquelle l'action peut être conçue en termes causaux et nomiques est bien entendu extrêmement controversée ; ici n'est pas le lieu de la défendre. J'ai essayé de le faire ailleurs (Kistler, à paraître, a).
8. Dans la terminologie de Kim (1998), il n'y a pas de *propriétés* de second ordre, mais seulement des *concepts* de second ordre et des *prédicats* de second ordre qui les désignent ou expriment. En revanche, dans la mesure où une propriété est réelle, elle est de premier ordre. Tout en approuvant ce raisonnement, je continuerai à parler de « propriétés de second ordre » : cela est possible à condition de donner un sens plus faible au terme de « propriété ». Dans ce sens faible, tous les prédicats expriment une propriété, même les prédicats négatifs et disjonctifs. On dira alors que ces propriétés ne sont pas des propriétés « réelles » : les propriétés réelles sont les propriétés de premier ordre qui constituent les termes de lois de la nature.
9. La distinction entre l'ordre logique et le niveau, dans l'ordre de la complexité, a été clairement faite par Kim (1997 ; 1998).

ou psychologique d'un organisme, et la réduction consiste à mettre en évidence l'existence d'une dépendance nomique entre cette propriété et des propriétés qui appartiennent à des parties du possesseur de la propriété macroscopique réduite. La réduction de la propriété d'être de l'hémoglobine peut servir d'illustration. L'hémoglobine est la substance qui réalise, chez les humains, la fonction de transporter l'oxygène des poumons vers les différents tissus du corps. Être une molécule d'hémoglobine est la propriété de premier ordre qui «réalise» le rôle caractérisé par la description fonctionnelle – de second ordre – d'être une propriété qui cause le transport d'oxygène dans le corps humain. Si dans d'autres organismes, d'autres molécules jouent ce rôle, alors le rôle du transporteur d'hémoglobine est multiréalisable au sens de Putnam et Kim. Mais l'hémoglobine est aussi «multiréalisable» dans le sens de la réduction entre *niveaux*: différentes molécules composées par différentes chaînes d'acides aminés donnent lieu à la conformation caractéristique de l'hémoglobine<sup>10</sup>. La composition de parties (acides aminés) *différentes* donne alors lieu à la *même* propriété du tout composé: la propriété d'être une molécule d'hémoglobine avec ses pouvoirs causaux caractéristiques. Cela explique pourquoi il est tout aussi légitime de parler de *l'hémoglobine* (au singulier) – en faisant référence à la molécule caractérisée par sa conformation globale et sa réactivité – et *des hémoglobines* (au pluriel) – en faisant référence aux différentes structures premières que sont les chaînes d'acides aminés dont la molécule est composée.

Fodor lui-même ne fait pas la distinction entre ces deux manières de concevoir l'ordre logique des propriétés «multiréalisables» (c'est-à-dire qui ont différentes propriétés réalisatrices). Cependant, sa stratégie de concevoir le rapport des propriétés de haut niveau aux propriétés de plus bas niveau à partir du modèle de la réduction interthéorique de Nagel mentionné plus haut montre que l'on peut les concevoir toutes les deux comme des propriétés de premier ordre. La réduction de la température à une propriété microphysique peut servir d'illustration.

Il est devenu possible de réduire la thermodynamique à la mécanique statistique lorsqu'on a découvert des principes de liaison tels que celui qui lie la propriété macroscopique de température ( $T$ ) à des propriétés microscopiques, telles que la masse d'une molécule du gaz et la vitesse des molécules  $v$  (c'est la moyenne des carrés des vitesses des molécules

10. Cf. Rosenberg (1985, p. 76).



qui intervient dans la formule de liaison). De manière analogue, la réduction d'une théorie psychologique passe par des principes de liaison qui expriment la corrélation nomique entre propriétés d'une science de niveau inférieur, par exemple la neurophysiologie, et propriétés de la science de haut niveau, c'est-à-dire dans ce cas la psychologie.

Selon l'une des interprétations des principes de liaison envisagées par Nagel, ces principes sont des lois de la nature, d'où l'appellation «lois-pont». On pourrait donc penser que l'existence de lois-pont entre une propriété  $S_1$  de haut niveau et des propriétés de plus bas niveau suffit pour garantir la réduction de  $S_1$ . Fodor cherche à montrer que cela n'est pas le cas : selon lui, les propriétés étudiées par les sciences spéciales ne peuvent pas être réduites bien qu'elles soient les termes de lois-ponts. La raison est que les propriétés qui font l'objet d'une science de haut niveau, ou «science spéciale» (par exemple la biologie), sont typiquement corrélées par des liaisons nomiques, non pas avec *une* propriété  $P$  de la science réductrice (par exemple la chimie), mais avec plusieurs, de sorte que le principe de liaison relie le prédicat  $S_1$  à un prédicat disjonctif<sup>11</sup> « $S_1x \leftrightarrow P_1x \vee P_2x \vee \dots \vee P_nx$ » (Fodor, 1974, p. 138, trad. p. 432).

Dans ce modèle, tous les prédicats sont de premier ordre. En particulier et contrairement à la conception fonctionnaliste, le prédicat de haut niveau  $S_1$  est conçu comme un prédicat désignant une propriété naturelle de premier ordre : en effet  $S_1$  constitue à la fois le terme d'une loi appartenant au niveau d'une science spéciale (omettant les quanteurs, Fodor écrit  $S_1x \rightarrow S_2x$ ), et le terme d'une loi-pont  $P_i \rightarrow S_1$  (sur la forme et la nature de laquelle nous reviendrons) pour chaque «réalisateur local»  $P_i$ .

Mais Fodor ne relève pas cette différence de conception entre son modèle du rapport entre propriétés appartenant à différents niveaux d'organisation et de la réalisabilité multiple, par rapport à la conception fonctionnaliste élaborée par Putnam. Au contraire, les rares exemples de propriétés multiréalisables qu'il offre, telles que la propriété d'être un échange monétaire, sont des propriétés fonctionnelles dont la réduction pourrait (en tout cas, dans un premier temps) prendre la forme d'une réduction

11. Fodor s'exprime en termes de *prédicats*, non de *propriétés* : le principe de liaison relie le prédicat  $S$  à un prédicat disjonctif. Sans préjuger le débat métaphysique sur la question de savoir s'il existe une propriété réelle pour chaque prédicat, je continue à m'exprimer en termes de propriétés. Il y a en effet un sens métaphysiquement innocent où il n'y a là que deux manières de parler, correspondant aux modes formel et matériel de Carnap (1931).



du rôle à l'occupant (c'est-à-dire une réduction entre propriétés appartenant à différents ordres logiques, comme dans les conceptions de Putnam et de Kim), au sens où différentes réalisations de la propriété fonctionnelle peuvent exercer la fonction : dans cet exemple,  $S_1$  désigne la propriété d'être un échange monétaire,  $P_1$  la propriété d'être un billet de dollar,  $P_2$  la propriété d'être un chèque signé, et les  $P_i$  toutes les autres incarnations matérielles possibles d'un échange économique. Cet exemple suggère l'interprétation selon laquelle  $S_1$  est un prédicat de second ordre et les  $P_i$  sont des propriétés de premier ordre, toutes appartenant au même niveau macroscopique. En ce sens, Fodor semble, comme Putnam, considérer comme équivalents les concepts de propriété multiréalisable (et donc irréductible) et celui de propriété fonctionnelle, les différentes propriétés réalisatrices étant conçues comme différentes propriétés pouvant jouer le rôle fonctionnel en question.

Cependant, d'autres passages de l'article de Fodor semblent compatibles avec la conception alternative selon laquelle tous les prédicats en jeu désignent des propriétés de premier ordre : lorsque Fodor affirme que « des événements psychologiques de même espèce sont parfois associés à des événements neurologiques d'espèces différentes » (Fodor, 1974, p. 137, trad. p. 430), il est possible d'interpréter l'événement psychologique comme une (instance de) propriété désignée par un prédicat  $S$  de premier ordre, et les événements neurologiques comme des (instances de) propriétés, elles aussi de premier ordre, mais appartenant à des parties cérébrales de l'organisme, c'est-à-dire à un *niveau* inférieur. « L'association » dont parle Fodor n'est alors plus celle entre un occupant (de premier ordre) de rôle fonctionnel par rapport à ce rôle (de second ordre), mais plutôt la relation qui caractérise les termes d'une « micro-réduction ». En effet, le but des neurosciences cognitives est de construire des explications réductrices qui permettent de comprendre l'existence des propriétés cognitives de l'organisme à partir des propriétés des parties neuronales de ces organismes et de leur interaction.

La conception fodorienne montre le chemin d'une conception du rapport entre propriétés psychologiques et neurologiques, selon laquelle ce sont des propriétés de même (premier) ordre logique qui font l'objet de sciences de différents *niveaux*. La thèse selon laquelle les propriétés psychologiques sont réalisables par différentes propriétés neurologiques signifie alors que différentes propriétés de niveau neurologique déter-

minent, en vertu de leurs interactions nomiques, la même propriété psychologique de l'organisme.

Ce changement de perspective a des conséquences majeures pour la question du matérialisme. Comme Putnam le fait remarquer lui-même, la conception fonctionnaliste des états mentaux ne garantit pas la vérité du matérialisme. De son hypothèse selon laquelle ce n'est que la structure fonctionnelle qui détermine les propriétés psychologiques, Putnam tire la conséquence que l'incarnation matérielle est accidentelle et ne contribue nullement à déterminer les propriétés psychologiques. Puis, comme il en a l'habitude, Putnam va jusqu'au bout de son raisonnement, en tirant explicitement la conséquence que la matérialité même des états qui occupent les rôles fonctionnels n'est pas requise : non seulement « nous pourrions être faits de fromage suisse, et cela ne changerait rien » (Putnam, 1975, p. 134) à notre esprit, mais nous pourrions avoir le même esprit sans avoir un cerveau matériel : nous pourrions avoir, à la place de nos cerveaux, des « âmes à la mode ancienne [*good old fashioned souls*] » (Putnam, 1975, p. 136), tout en ayant le même esprit que dans le monde actuel où les états du cerveau occupent les rôles qui déterminent nos états mentaux.

Cela fait-il une différence que les gens pourvus d'une âme ont pour ainsi dire des cerveaux immatériels, et que les gens pourvus d'un cerveau ont des âmes matérielles ? Ce qui compte c'est la structure commune [...] et non le hardware, fut-il éthéré [*not the hardware, be it ever so ethereal*] (Putnam, 1975, p. 136)<sup>12</sup>.

La conception fodorienne montre le chemin vers une position matérialiste. Cependant, la forme particulière que Fodor lui-même donne à sa conception des propriétés psychologiques ne garantit pas davantage leur matérialité que la conception putnamienne. Car comme nous allons le voir, les principes de liaison entre propriétés neurologiques et psychologiques ne garantissent pas l'homogénéité ontologique de leurs termes : le fait que les propriétés neurologiques soient matérielles, et le fait qu'elles déterminent nomiquement les propriétés psychologiques, ne garantit pas que ces dernières soient matérielles elles aussi.

12. Putnam tire cette conséquence dès le célèbre article « La nature des états mentaux » (Putnam, 1967) dans lequel il introduit l'hypothèse fonctionnaliste selon laquelle un état mental tel que « l'état d'éprouver de la douleur [...] est un état fonctionnel d'un organisme tout entier » (Putnam, 1967, p. 154, trad. p. 276). Il y constate en effet que « l'hypothèse de l'état fonctionnel n'est *pas* incompatible avec le dualisme ! » (Putnam, 1967, p. 157, trad. p. 280, italiques de Putnam), car l'hypothèse fonctionnaliste conçoit les états mentaux comme des états porteurs d'information qui « ne sont, bien sûr, ni mentaux ni physiques *per se* » (Putnam, 1967, p. 157, trad. p. 280, italiques de Putnam).

## 2 – Les propriétés mentales comme survenantes sur les propriétés du cerveau

Afin de concevoir le rapport entre le mental et physique d'une manière qui permet de concilier la vérité du physicalisme avec l'irréductibilité du mental au physique, Donald Davidson a suggéré d'utiliser le concept de survenance<sup>13</sup>:

[...] les caractéristiques mentales sont en un certain sens dépendantes des caractéristiques physiques, ou survenantes par rapport à elles. On peut interpréter cette survenance comme signifiant qu'il ne peut pas y avoir deux événements qui soient semblables sous tous leurs aspects physiques mais qui diffèrent sous un aspect mental quelconque. Ce genre de dépendance ou de survenance n'implique pas la réductibilité par l'intermédiaire de lois ou de définitions (Davidson, 1980, p. 214, trad. p. 248).

L'idée fondamentale de la survenance est que, si un ensemble de propriétés (par exemple psychologiques) P survient sur un autre ensemble de propriétés (par exemple neurologiques) N, alors il est impossible que deux objets qui partagent toutes les propriétés de type N diffèrent néanmoins à l'égard de leurs propriétés de type P. Ce concept a été décliné en de nombreuses variantes dont l'adéquation au rapport entre le mental et le physique a donné lieu à un débat riche et controversé<sup>14</sup>. En prenant le risque de conclure prématurément un débat encore ouvert, j'avance l'hypothèse selon laquelle la survenance forte (locale) est le concept le plus prometteur pour rendre compte du rapport entre propriétés mentales et physiques. Un ensemble de propriétés P survient fortement sur un ensemble de propriétés N si et seulement si, nécessairement, pour toute propriété F dans P, si un objet possède F, alors il existe une propriété G dans N telle que cet objet possède G, et nécessairement tout ce qui possède G possède F<sup>15</sup>. Selon une autre formulation, les propriétés de l'ensemble P surviennent fortement sur les propriétés de l'ensemble N si et seulement si, nécessairement toute paire d'objets (que les objets de cette paire se trouvent dans le même monde ou dans différents mondes possibles) dont les objets ne diffèrent pas à l'égard de leurs propriétés N ne diffèrent pas à l'égard de leurs propriétés P; ou « plus simplement,

13. Davidson (1980) donne un argument assez différent de ceux de Putnam et Fodor pour l'irréductibilité de l'esprit. L'espace nous manque pour le présenter ici. Cf. Engel (1994).

14. Les travaux les plus importants de Kim sur les différents espèces de survenance ont été rassemblés dans Kim (1993). Les contributions rassemblées dans Savellos et Yalçin (1995) montrent l'importance de ce concept pour de nombreux problèmes philosophiques.

15. Cf. Kim (1997a, p. 188).

l'indiscernabilité à l'égard de N entraîne l'indiscernabilité à l'égard de P» (Kim, 1997a, p. 188, lettres modifiées).

La thèse selon laquelle le mental *survient*, au sens que l'on vient d'indiquer, sur le physique a pu sembler fonder une position de compromis, compatible à la fois avec l'irréductibilité du mental au physique – dont les arguments de Putnam, Fodor et Davidson avaient convaincu de nombreux philosophes – et avec le physicalisme «des occurrences». Selon cette doctrine, tous les événements mentaux, c'est-à-dire les événements qui ont des propriétés mentales, sont aussi des événements physiques, au sens d'avoir des propriétés physiques, et en ce sens, tous les événements sont des événements physiques. La thèse de la survenance semblait présenter l'avantage de laisser ouverte la question de la nature des propriétés mentales, et en particulier celui d'être compatible avec leur irréductibilité supposée, tout en garantissant leur ancrage dans le physique.

Or il s'est avéré que la survenance est une relation trop faible pour garantir le matérialisme. Horgan (1993) et Kim (1993a) montrent que la survenance est une simple relation de corrélation entre deux ensembles de propriétés qui est «superficielle» en ce sens qu'elle n'indique pas la raison de cette corrélation. Au lieu de trancher la question de savoir pourquoi certaines propriétés physiques sont systématiquement accompagnées de certaines propriétés mentales, soit en un sens dualiste en affirmant comme Malebranche et Leibniz qu'elle est due à l'intervention divine, soit en un sens matérialiste en disant par exemple que ces propriétés sont identiques, la thèse de la survenance se contente d'affirmer la corrélation, sans prétendre l'expliquer.

Lorsque nous considérons la simple affirmation de la survenance de l'esprit sur le corps, et la comparons avec ces options traditionnelles, nous sommes frappés par son échec de remplir cette tâche explicative. Car elle affirme simplement une relation de dépendance d'une espèce non spécifique, sans rien faire de plus pour expliquer la nature de la corrélation psychophysique. Mais la survenance elle-même n'est pas une relation explicative. Ce n'est pas une relation métaphysique «profonde»; c'est plutôt une relation de «surface». [...] Nous n'avons aucune théorie sur le rapport entre corps et esprit, tant que nous n'avons rien à dire sur le fondement de la corrélation entre les propriétés mentales et physiques (Kim, 1993, p. 167)<sup>16</sup>.

16. Cf. aussi Kim (1990, p. 26-27, repr. 1993, p. 159). Charles (1992) et McLaughlin (1995) ont également insisté sur la faiblesse de la relation de survenance qui n'affirme que la covariation systématique des propriétés des deux ensembles. Charles explique que la survenance ne contient ni la relation de priorité explicative ni la relation asymétrique selon laquelle la base de survenance est la *base ontologique* des propriétés survenantes. Il fait par exemple remarquer que dans un monde déterministe où S a toujours l'effet T, et où T ne peut être

Au lieu de s'engager en faveur d'une conception matérialiste du rapport entre corps et esprit, l'avocat de la survenance évite de prendre position sur l'origine de la survenance. Le prix à payer pour cette neutralité métaphysique de la seule affirmation de la survenance forte des propriétés mentales sur les propriétés physiques, est que la position qui en résulte est compatible avec le parallélisme ou l'occasionalisme : dans la mesure où l'intervention divine a une force modale, c'est-à-dire garantit la corrélation entre propriétés physiques et mentales dans tous les mondes possibles, elle garantit la survenance forte. Comme le remarque Kim, la survenance forte est même compatible avec une version du dualisme des substances qui soutient que, malgré le fait que les âmes soient ontologiquement séparées des corps matériels, «ce qui advient aux âmes est entièrement déterminé, une fois qu'il est déterminé ce qui advient dans le domaine matériel» (Kim, 1988, p. 133, repr. 1993, p. 114)<sup>17</sup>.

L'existence d'une corrélation universelle entre propriétés mentales et propriétés physiques, et même une corrélation nécessaire comme dans la survenance forte, ne contient aucune indication quant à l'origine ou à l'explication de cette corrélation. En particulier, elle n'implique ni que les propriétés mentales *dépendent* des propriétés physiques<sup>18</sup> ni que les

---

causé que par S, toutes les propriétés qui surviennent sur S, surviennent aussi sur T. Cela montre que la survenance ne permet pas d'identifier «la base appropriée de l'occurrence d'une propriété mentale» (Charles, 1992, p. 275). Comme McLaughlin le fait remarquer, même la survenance forte «n'implique pas des connexions explicatives entre propriétés survenantes et subvenantes» et «si la réduction est une relation *explicative*, alors la SSm [la survenance forte, telle qu'elle est définie avec l'opérateur modal de nécessité, dans McLaughlin, 1995, p. 25] des propriétés A sur les propriétés B avec nécessité métaphysique ne suffit par pour la réduction» (McLaughlin, 1995, p. 48).

17. Ailleurs, Kim explique que la thèse de la survenance est compatible avec «une foule de positions classiques sur le problème du rapport entre corps et esprit. Elle constitue en fait une implication commune de nombreuses théories sur le rapport du corps à l'esprit qui s'excluent mutuellement» (Kim, 1997a, p. 189). Entre autres, il y mentionne l'émergentisme selon lequel le physique détermine le mental par des lois primitives et inexplicables : c'est selon Kim «une forme de dualisme qui considère les propriétés mentales comme des pouvoirs causaux intrinsèques non-physiques» (Kim, 1997a, p. 189). L'épiphénoménalisme qui soutient que le mental est causé par le physique sans lui-même rien causer, implique également la survenance.

18. Je conteste ici le jugement de Kim (exprimé dans la longue citation de la page précédente, de Kim, 1993, p. 167) qui affirme qu'il s'agit d'une forme de dépendance quoiqu'elle soit non spécifique. Yablo aussi constate que les définitions de la survenance (même forte) ne garantissent que la *corrélation* des propriétés concernées, mais non l'existence d'une quelconque *dépendance* qui fonde cette corrélation. «Les définitions habituelles de la survenance en termes de corrélations ne parvenaient pas à rendre compte d'une telle dépendance [*failed to capture any such dependency*]» (Yablo, 1997, p. 275). Toutefois, il ne s'agit peut-être que d'un désaccord terminologique. Après tout, Kim lui-même s'exprime ailleurs d'une manière qui suggère que la survenance ne suffit pas pour l'existence d'une dépendance : «Tout physicien qui croit dans la réalité du mental doit accepter l'existence d'une corrélation psychophysique universelle sur le plan des propriétés... *plus* l'affirmation selon laquelle il existe une relation de dépendance sous-jacente à cette corrélation» (Kim, 1993, p. 169).

propriétés physiques *déterminent* les propriétés mentales : dans les doctrines parallélistes et occasionalistes, les instances des propriétés de deux types sont déterminées par une cause indépendante, et ne dépendent que d'elle, à savoir la volonté de Dieu. Cela montre deux choses : 1) étant donné que la corrélation affirmée par la survenance peut avoir différentes causes, l'affirmation de la survenance ne constitue pas une explication complète du rapport entre les propriétés survenantes et les propriétés dans leur base de survenance<sup>19</sup>. 2) L'affirmation de la survenance entre propriétés mentales et physiques ne garantit pas une position matérialiste. Comme le dit Horgan, « la seule survenance des propriétés et faits d'ordre supérieur sur des propriétés et faits physiques ne peut pas être suffisante pour conférer de la respectabilité matérialiste » (Horgan, 1993, p. 565). Ce qui rendrait l'explication du rapport complet, tout en montrant son caractère matérialiste, ce serait la mise en évidence d'une relation plus forte que la survenance (Horgan baptise cette relation « plus-que-survenance » [*superdupervenience*]) : une forme de « survenance ontologique qui peut être expliquée de manière robuste et matérialiste » (Horgan, 1993, p. 566).

Selon la conception que je voudrais suggérer, les propriétés psychologiques sont déterminées par des lois de composition, elles-mêmes fondées sur des lois d'interaction, à partir des propriétés des parties des organismes qui les possèdent. Cette « détermination nomique » est censée offrir le soubassement métaphysique que Horgan appelle de ses vœux : elle vise à identifier la nature de la détermination des propriétés P d'un objet complexe *o* par les propriétés des composantes de *o* et de leurs relations. Cela permet de comprendre *la raison* pour laquelle les propriétés P surviennent sur les propriétés des composantes, survenance qui constitue souvent le point de départ de la réflexion philosophique sur le rapport entre les propriétés mentales et les propriétés physiques sous-jacentes (par exemple dans Kim, 1998). La détermination nomique implique donc la survenance, mais va bien au delà d'elle en proposant une explication de son origine<sup>20</sup>. L'explication selon laquelle les propriétés mentales

19. Cf. Kim (1990, p. 26-27), Horgan (1984), Horgan (1993, p. 577 sq.).

20. La détermination nomique est aussi censée offrir un cadre pour la conception des propriétés mentales comme *émergentes et malgré cela réductibles*. Nous devons nous contenter ici de quelques brèves remarques à ce sujet. Le fait que l'émergence soit compatible avec la survenance a clairement été exprimé par Broad (1925, p. 67-68). Nous pourrions exprimer le rapport entre survenance et émergence, en modifiant la formule de Blackburn qui dit que « la survenance est l'émergence en tant que déterminée physiquement [*supervenience is physically fixed emergence*] » (Blackburn, 1993, p. 233). Selon la conception développée

sont naturellement déterminées par les propriétés des parties de l'organisme, notamment par l'activité de ses neurones, et par les interactions entre ces parties, en vertu de lois de composition, exclut notamment le dualisme des substances, ainsi que le parallélisme et l'occasionalisme.

### 3 – Réduction locale

L'argument pour l'irréductibilité des propriétés psychologiques aux propriétés physiques a récemment été attaqué de deux manières indépendantes. D'un côté, Bechtel et Mundale (1999) montrent que nombre de découvertes récentes concernant le soubassement neuronal de différentes capacités cognitives, notamment la corrélation entre l'état d'activation de neurones individuels avec certains états mentaux, ainsi que l'existence et le rôle fonctionnel de différentes aires cérébrales et des chemins de transmission d'information entre différentes aires cérébrales, n'ont été possibles que sur la base de l'hypothèse de travail selon laquelle l'organisation fonctionnelle de notre appareil cognitif *n'est pas* multi-réalisable, dans la mesure où ces découvertes reposent sur des recherches faites sur les cerveaux de différentes espèces animales et sur le cerveau humain. La découverte de l'architecture cérébrale sous-jacente à l'architecture cognitive, par exemple, dans le cas du processus de traitement de l'information visuelle, «présuppose qu'il existe une réalisation commune à différentes espèces, des mécanismes du traitement de l'information visuelle» (Bechtel et Mundale, 1999, p. 201). Il n'y a qu'en niant la réalisabilité multiple que les neurosciences peuvent être utiles pour «guider la décomposition et la compréhension des systèmes cognitifs» (Bechtel et Mundale, 1999, p. 204).

---

pée ici, les propriétés émergentes sont déterminées par des lois qui ne restent pas nécessairement des «*nomological danglers*», c'est-à-dire qui sont promises à devenir des théorèmes dérivables à partir des lois d'interaction des propriétés de la base d'émergence. En ce sens, toute propriété émergente est «physiquement fixée (déterminée)». (Notre conception selon laquelle le caractère émergent d'une propriété est compatible avec sa réductibilité, contredit la conception traditionnelle de l'émergence comme relation qui présuppose l'irréductibilité. Ici n'est pas le lieu pour développer une conception nouvelle de l'émergence qui ne présuppose plus l'irréductibilité. Pour quelques suggestions prometteuses en ce sens, cf. Batterman, 2000 et Rueger, 2000.) Mais la relation d'émergence est plus forte que celle de survenance; elle l'implique alors que la survenance est compatible avec des lois de liaison irréductibles, et même avec la corrélation extensionnelle brute sans loi. Par conséquent, l'émergence en tant que déterminée physiquement implique la survenance, mais la survenance n'implique pas l'émergence en tant que déterminée physiquement. Contrairement à ce que soutient Blackburn, une propriété émergente est plus étroitement liée à ses propriétés de base qu'une propriété survenante ne l'est aux propriétés de sa base.

L'argument « transcendantal » de Bechtel et Mundale contre la multiréalisabilité est certes contestable : selon eux, aucune réduction possible du mental au neurologique ne peut conduire à la découverte de réalisations multiples parce que la méthode qui permet aux neurosciences cognitives de découvrir les réductions psychophysiques *présuppose* que les capacités cognitives sont réalisées de *la même façon* dans les *différentes espèces* animales étudiées. L'exemple suivant montre que cette présupposition n'empêche pas la découverte que les capacités cognitives sont multiréalisables. (Il empêche seulement la possibilité de découvrir des cas de multiréalisabilité *radicale* où les différentes propriétés réalisatrices n'ont rien en commun.) En revanche, cet exemple constitue au moins une raison inductive pour penser que Bechtel et Mundale ont raison de dire que d'importantes capacités cognitives sont partagées et non radicalement multiréalisables.

Les recherches sur le mécanisme moléculaire et cellulaire sous-jacent à une capacité cognitive importante, à savoir celle de la fixation de la mémoire à long-terme, ont montré que ce mécanisme est partagé par des espèces animales extrêmement différentes : les processus cellulaires et moléculaires qui permettent au cerveau d'effectuer les modifications synaptiques qui sont le support de la mémoire à long terme sont essentiellement les mêmes chez *Drosophila*, chez l'aplysie (*Aplysia californica*), chez la souris et chez l'homme. Squire et Kandel parlent du :

conservatisme évolutionnaire qui caractérise les mécanismes moléculaires sous-jacents aux processus mentaux. [...] Ces mécanismes communs ont été conservés à travers l'histoire évolutionnaire des espèces : on les trouve aussi bien chez des invertébrés aussi simples que *Drosophila* et *Aplysia* que chez des mammifères aussi complexes que les souris (Squire et Kandel, 1999, p. 155).

Cependant, la présence de traits communs possédés par les différents mécanismes qui assurent, dans les différentes espèces, la consolidation de la mémoire à long terme, ne justifie pas la conclusion, tirée notamment par Bickle (2003, p. 132, 141 ff.), selon laquelle cette consolidation n'est pas réalisée de différentes manières dans les différentes espèces. Dans la présentation détaillée des mécanismes sous-jacents à la consolidation de la mémoire dans différentes espèces qu'offre Bickle, on trouve en effet, à côté des traits communs, de nombreuses différences : par exemple, la comparaison des protéines CREB qui jouent un rôle déterminant dans la transcription des gènes conduisant à la modification des synapses suite à l'apprentissage, montre que :



l'un de ces produits [présent chez *Aplysia*], le polypeptide isoforme CREB1a, montre une homologie à l'égard de 95 % de sa séquence d'acides aminés par rapport aux protéines CREB des mammifères, ce qui signifie que 19 sur 20 acides aminés dans les séquences de ces protéines étaient identiques à travers ces espèces très différentes (Bickle, 2003, p. 144).

Les molécules qui jouent le rôle de la protéine CREB diffèrent peu ; cependant, même avec 5 % de différences, il s'agit bien de réalisation multiple. Quelques pages plus loin, un autre facteur de transcription illustre le même fait : la forme d'une protéine (ApC/EBP) présente dans le mécanisme de la fixation de la mémoire à long terme d'*Aplysia* a des traits communs mais aussi des traits divergents, par rapport à la forme de la protéine présente dans le mécanisme homologue du rat (Bickle, 2003, p. 147). Bickle a donc tort d'affirmer que ces découvertes réfutent l'hypothèse de la réalisabilité multiple ; elles nous avertissent seulement du fait que les arguments purement conceptuels de Fodor et Putnam suffisent à établir tout au plus qu'il est *concevable* que les capacités cognitives soient réalisées par des mécanismes radicalement différents, mais non que les mécanismes qui les réalisent dans les différentes espèces animales terrestres sont réellement radicalement différents.

L'autre attaque contre l'irréductibilité des propriétés mentales part d'un réexamen de la logique de la réduction, dans le cadre du modèle proposé par Nagel. Comme nous l'avons vu, dans ce modèle, la réduction d'une théorie de haut niveau passe par l'hypothèse de l'existence de principes de liaison qui expriment la corrélation nomique des propriétés de la science réductrice, par exemple la neurophysiologie, avec les propriétés de la science réduite, par exemple la psychologie. Selon l'une des interprétations envisagées par Nagel de ces principes de liaison, ces derniers sont des lois de la nature. Comme nous l'avons vu aussi, l'argument anti-réductionniste, notamment dans la forme que lui donne Fodor, présuppose que ces lois aient une forme biconditionnelle. Selon cet argument, dans le cas d'une propriété psychologique qui peut être réalisée, sur le plan neurophysiologique, par un ensemble en principe infini de structures, il ne peut pas y avoir de telle loi puisque son terme neurophysiologique serait une disjonction ouverte, ayant un nombre infini de termes. Or, la question de savoir si les principes de liaison doivent avoir une forme *biconditionnelle*, ou si une réduction peut être accomplie sur la seule base de principes de liaison de forme conditionnelle, est controversée. La réponse dépend de la réponse à une autre question : la réduc-

tion requiert-elle la *dérivabilité* des lois de la théorie réduite, ou peut-on au contraire se satisfaire d'une relation moins forte de «*connectabilité*», dans laquelle on déduit de la théorie réductrice, non pas la loi de la théorie réduite elle-même, mais son «image» qui partage avec elle certaines propriétés structurelles importantes. Nagel lui-même note qu'il y a des cas où la dépendance entre un prédicat A de la théorie réduite, par rapport à un prédicat B de la théorie réductrice, n'est pas réciproque et où, par conséquent, les principes pont n'ont pas de forme biconditionnelle, mais seulement conditionnelle :

Le lien entre A et B n'a pas nécessairement une forme biconditionnelle, et peut par exemple n'être qu'un conditionnel à sens unique : si B alors A (Nagel, 1961, p. 355, note).

De telles lois-pont de forme conditionnelle suffisent en effet à garantir la connexion entre les deux théories ; leur existence satisfait ce que Nagel appelle «la condition de connectabilité» (Nagel, 1961, p. 354). Comme le note Richardson :

Il n'est pas nécessaire que les «relations appropriées» requises par la condition de connectabilité soient de forme biconditionnelle (Richardson, 1979, p. 548).

Mais Nagel requiert davantage d'une réduction que la seule connectabilité : la seconde condition, appelée «condition de dérivabilité» exige que la réduction permette de déduire toutes les lois de la théorie réduite de la théorie réductrice :

Il faut que toutes les lois de la science secondaire [c'est-à-dire réduite], y compris celles qui contiennent le terme «A», soient logiquement dérivables des prémisses théoriques et les définitions coordinatrices qui leur sont associées [c'est l'interprétation des principes pont retenue par Nagel] dans la discipline première [c'est-à-dire la théorie réductrice] (Nagel, 1961, p. 354).

Le point crucial est le suivant : Les principes de liaison de forme conditionnelle, de type  $P_1x \rightarrow Sx$ ,  $P_2x \rightarrow Sx$ , ne suffisent pas à déduire logiquement les lois de la théorie réduite à partir des lois de la (ou des) théorie(s) réductrice(s). Mettons que les lois des théories réductrices ont la forme  $P_1x \rightarrow P_1^*x$ ,  $P_2x \rightarrow P_2^*x$ , etc., et les lois-pont ont la forme conditionnelle :  $P_1x \rightarrow Sx$ ,  $P_2x \rightarrow Sx$ ,  $P_1^*x \rightarrow S^*x$ ,  $P_2^*x \rightarrow S^*x$ , etc. Pour déduire logiquement la loi  $Sx \rightarrow S^*x$ , une loi de la forme  $Sx \rightarrow P_1x$  est indispensable, c'est-à-dire une loi exprimant le fait que l'exemplification de l'antécédent S de la loi à réduire est suffisant pour l'exemplification d'une propriété désignée par l'antécédent d'une loi de la science

réductrice. S'il existe une telle loi «du haut vers le bas», par exemple  $Sx \rightarrow P_5x$ , alors on peut déduire, par transitivité, de  $Sx \rightarrow P_5x$ , de la loi réductrice pertinente,  $P_5x \rightarrow P_5^*x$ , et de la loi-pont pertinente pour  $P_5^*$ ,  $P_5^*x \rightarrow S^*x$ , que:  $Sx \rightarrow S^*x$ . Mais la loi «du haut vers le bas»  $Sx \rightarrow P_5x$  est précisément ce qui fait défaut dans les cas de réalisabilité multiple. En l'absence d'une telle loi, la dérivation purement logique de la loi de haut niveau à partir de lois de bas niveau, et de lois-pont conditionnelles «du bas vers le haut» est impossible. Nagel le dit très clairement: au cas où les seules lois-pont disponibles sont de forme conditionnelle «si B, alors A [...] "A" n'est pas remplaçable par "B", et donc la science secondaire ne sera en général pas déductible d'une théorie de la discipline primaire» (Nagel, 1961, p. 355, note).

Richardson a raison de noter que des lois-pont de forme conditionnelle suffisent à la connectabilité, mais il a tort de dire que de telles lois suffisent aussi à la dérivabilité, et donc à la réduction au sens le plus fort<sup>21</sup>:

La seule découverte de conditions suffisantes à un niveau inférieur d'organisation aptes à rendre compte de phénomènes que l'on avait à l'origine expliqués à un niveau plus élevé, permet de rendre compte, de manière adéquate, de la dérivabilité, avec sa parcimonie explicative; et cela aussi ne requiert pas davantage qu'une correspondance [*mapping*] de types de niveau inférieur à des types de niveau supérieur et non pas une correspondance des types supérieurs aux types inférieurs (Richardson, 1979, p. 548).

Cependant, il me semble néanmoins que Richardson exprime une vérité importante: lorsque nous découvrons la propriété microscopique, par exemple la propriété neurophysiologique qui est sous-jacente à une propriété biologique donnée, ainsi que le mécanisme microscopique qui donne lieu à la propriété biologique macroscopique, nous considérons que cette découverte permet de donner une explication réductrice du phénomène, même si, dans d'autres organismes, la propriété microscopique et/ou le mécanisme est différent. Prenons l'exemple de l'explication réductrice du «signal» contenu dans une protéine qui permet à la molécule d'atteindre sa destination au sein de la cellule<sup>22</sup>. Selon «l'hypothèse du signal» chaque protéine qui est synthétisée dans le cytoplasme par des complexes de ribosomes et d'ARN possède une propriété qui détermine son cheminement vers sa destination fonctionnelle: cette propriété est le signal permettant l'orientation des protéines. Or il apparaît que des

21. Ce point est clairement exprimé par Marras (2002, p. 248 sq.).

22. Cf. Kincaid (1990).

propriétés microscopiques très différentes peuvent «jouer le rôle» de signal dans différents organismes, mais aussi pour différentes protéines dans un organisme donné. À titre d'exemple,

les séquences [d'acides aminés à l'intérieur de la protéine] qui signalent l'insertion de la protéine dans le réticulum endoplasmique [...] peuvent avoir des longueurs qui diffèrent du simple au triple, interagissent apparemment chimiquement et physiquement de différentes manières avec les lipides de la membrane, peuvent ou non être découpées pendant l'exécution de leur fonction [...] et sont parfois particulières à l'espèce [...] (Kincaid, 1990, p. 581).

La découverte de chacun des mécanismes permettant à une micropropriété sous-jacente à la propriété signalante d'une protéine d'orienter la protéine vers sa destination constitue bien une explication réductrice. En ce sens, il est correct de dire, avec Richardson, que l'existence de «mécanismes (actuels ou possibles) alternatifs» à un mécanisme donné «n'empêche pas la réduction» (Richardson, 1979, p. 549). Mais il faut être conscient du fait que la réduction accomplie par la découverte d'une micropropriété et du mécanisme microscopique associé, n'est pas une réduction au sens de la dérivabilité logique de la loi de haut niveau, au sens de Nagel. La contradiction entre l'affirmation de Richardson que nous venons de citer avec la thèse de Kincaid selon laquelle la «diversité biochimique sous-jacente à l'unité biologique constitue l'obstacle fondamental à la réduction» (Kincaid, 1990, p. 583), n'est donc qu'apparente: Kincaid parle de la réduction au sens de dérivation logique, alors que la thèse de Richardson peut se défendre dès lors que l'on donne à «réduction» un sens plus faible<sup>23</sup>. Dans la mesure où il est suffisamment éclairant de dériver, pour l'un ou l'autre des types de systèmes possédant la propriété réduite, une loi structurellement équivalente à la loi réduite  $Sx \rightarrow S^*x$ , de forme  $P_i x \rightarrow P_i^* x$ , à partir de la théorie réductrice qui décrit correctement tel type de système possédant S, on peut considérer que cette découverte constitue une «explication réductrice» quoique non au sens nagélien de dérivabilité. En ce sens, Marras observe qu'il est légitime de dire que des propriétés biologiques qui sont réalisées de différentes façons dans différents types d'organismes peuvent néanmoins être réduites à condition d'affaiblir les conditions nagéliennes de la réductibilité, à savoir «si nous considérons que le noyau essentiel de la réduction n'est pas la dérivation des lois *actuelles* de la théorie réduite des lois

23. Richardson (1979) lui-même ne semble pas en être conscient: il prétend à tort que les lois-pont conditionnelles suffisent à la dérivabilité nagélienne.

de la théorie réductrice, mais seulement la dérivation des *images* de ces lois, dans des circonstances appropriées» (Marras, 2002, p. 249).

Dans le cadre d'une telle conception affaiblie de la réduction<sup>24</sup>, le fait qu'un nombre indéterminé et peut-être infini de propriétés physiques  $P_i$  peut être sous-jacent à la propriété mentale  $S$  n'empêche pas la réduction de  $S$ . La réalisabilité multiple est compatible avec la possibilité de réduire la propriété multiréalisable à ses différentes propriétés réalisatrices. Ce changement de perspective a plusieurs conséquences importantes. Premièrement, comme Kim (1992) l'a fait remarquer, de telles réductions ne sont que «locales» :

Si chacune des espèces psychologiques qui sont postulées dans une théorie psychologique a une réalisation physique pour une espèce donnée, la théorie peut être «localement réduite» à la théorie physique de cette espèce (Kim, 1992, p. 19, repr. 1993, p. 328).

Deuxièmement, le concept de réduction locale nous force à abandonner la thèse (défendue par Causey, 1977 et largement acceptée, entre autres par Fodor, 1974) selon laquelle la découverte d'une réduction prend nécessairement la forme de la découverte de *l'identité* de propriétés. Le fait qu'il n'y ait pas de loi de forme biconditionnelle qui fonde la réduction, constitue au contraire une raison de nier qu'il existe une propriété (unique) identique à la propriété réduite.

La prise de conscience de la possibilité de réductions locales constitue certainement un progrès important en direction d'un ancrage de l'esprit dans la matière, en montrant que la multiréalisabilité des propriétés psychologiques n'y oppose pas d'obstacle insurmontable. Cependant, tout comme la thèse de la survenance, la réductibilité locale à partir de lois-pont de forme conditionnelle et de la dérivation de lois «analogues» ne suffit pas à elle seule à garantir le matérialisme. Schaffner, Hooker et Churchland ont développé une conception plus faible de la réduction qui ne requiert pas la déduction de la théorie réduite à partir de la théorie réductrice, mais tient compte du fait que la plupart des réductions s'accompagnent de corrections apportées à la théorie réduite. Par conséquent, ce qui est déduit à partir de la théorie réductrice n'est pas l'ancienne théorie qui fait l'objet de la réduction, mais une nouvelle théorie qui lui ressemble: dans les termes de Schaffner (1967), la théorie  $T_2^*$  dérivée de

24. Schaffner (1967; 1993) a élaboré un tel modèle de la réduction de manière détaillé. Cf. aussi Hooker (1981), Churchland (1985) et Bickle (1998).

$T_1$  (la théorie réductrice) doit être dans une relation de «similarité étroite» avec la théorie originale qu'il s'agissait de réduire,  $T_2$ ; les prédictions numériques faites à partir de  $T_2^*$  doivent être «très proches» de celles faites à partir de  $T_2$ ; entre la théorie à réduire  $T_2$  et la théorie effectivement dérivable de  $T_1$ ,  $T_2^*$ , il doit y avoir une «analogie forte» ou «analogie positive» (Schaffner, 1967, p. 144)<sup>25</sup>. L'abandon de l'exigence de déductibilité et son remplacement par l'exigence de la déductibilité d'une théorie *analogue* à la théorie réduite, rend la conception de Schaffner, Hooker et Churchland en particulier compatible avec la multiréalisabilité: dans le cas d'une propriété multiréalisable, il existe plusieurs théories réductrices, chacune servant de base à la déduction d'une théorie analogue, sans que les différentes théories ainsi obtenues soient identiques entre elles.

La thèse selon laquelle la réductibilité locale ne suffit pas à garantir le matérialisme peut paraître surprenante à première vue, car selon une acception courante du terme «réductionnisme», une position réductionniste à l'égard d'un certain domaine ne reconnaît pas d'existence autonome à ce domaine. En ce sens, on considère par exemple, qu'un réductionniste par rapport à l'esprit soutient que l'esprit n'est en réalité «rien d'autre que le corps». Pour dissoudre l'apparence de paradoxe, il suffit d'observer que la conception nagelienne de la réduction partage avec la conception plus faible de Schaffner, Hooker et Churchland (qui, comme nous l'avons vu, permet de concevoir des propriétés multiréalisables comme néanmoins réductibles) la thèse fondamentale selon laquelle la connexion entre théorie réductrice et réduite (et donc indirectement entre propriétés de leurs niveaux respectifs) dépend de la découverte de lois de la nature. Or, de telles lois qui établissent l'existence d'une *dépendance* entre les propriétés de niveau inférieur désignées par les prédicats de la théorie (ou des théories) réductrice(s), ne garantissent pas que les propriétés qui constituent leurs termes soient *de même type*, ou qu'elles se

25. Dans son développement de la conception de Schaffner, Hooker (1981) soutient que l'existence d'une «relation d'analogie» entre la théorie  $T_R^*$  dérivée de la théorie de base  $T_B$  (comme «théorie de base») et la théorie qu'il s'agit de réduire,  $T_R$ , «autorise que l'on affirme qu'il existe (une sorte) de relation réductrice entre  $T_R$  et  $T_B$ » (Hooker, 1981, p. 49). Dans les termes de Churchland (1985), «une réduction consiste dans la déduction, à l'intérieur de  $T_N$  [comme «théorie nouvelle»; ce qui correspond à la théorie réductrice] non pas de  $T_O$  [comme «théorie ancienne (*old theory*)», ce qui correspond à la théorie réduite] elle-même, mais d'une image approximativement équivalente de  $T_O$ , une image encore exprimée dans le vocabulaire propre à  $T_N$ » (Churchland, 1985, p. 10). Bickle (1998) contient un exposé détaillé de la conception de la réduction développée par Schaffner, Hooker et Churchland.

ressemblent de quelque manière que ce soit. L'existence d'une loi de forme conditionnelle  $P_i x \rightarrow Sx$  ou biconditionnelle  $Px \leftrightarrow Sx$  qui joue le rôle de loi-pont et connecte les niveaux des théories réductrice et réduite ne signifie rien d'autre que le fait que l'exemplification de la propriété mentionnée dans l'antécédent ( $P$  ou  $P_i$ ) garantit, avec la nécessité nomologique propre aux lois, l'exemplification de la propriété mentionnée dans le conséquent ( $S$ ). Même dans le cas d'une loi de forme biconditionnelle, la loi n'exprime pas plus que le fait que les exemplifications des propriétés qu'elle relie sont corrélées, et qu'elles doivent l'être, où le mot «doivent» indique la force modale de la nécessité nomologique.

Il apparaît donc que le réductionnisme ne garantit le matérialisme ni dans la variante «forte» nagelienne, ni dans la forme affaiblie du modèle de Schaffner, Hooker et Churchland : une propriété  $S$ , et les lois de forme  $Sx \rightarrow S^*x$  auxquelles elle obéit, peut satisfaire les conditions de connectabilité et de dérivabilité (Nagel) ou de dérivabilité d'une image analogue (Schaffner-Hooker-Churchland), sans être matérielle : il suffit que les lois-pont l'associent avec la force modale de la nécessité nomologique à ses propriétés réductrices matérielles, de sorte que les lois du niveau de  $S$  (ou des lois structurellement analogues) puissent être dérivées à partir du niveau réducteur. Or ces contraintes purement formelles laissent la nature ontologique des propriétés réduites entièrement indéterminée. Comme le note Fodor, la découverte d'une réduction selon le modèle nagelien est compatible avec «une sorte de dualisme psychophysique non-cartésien : un dualisme d'événements et/ou de propriétés plutôt que de substances» (Fodor, 1974, p. 130, trad. p. 421)<sup>26</sup>. La réduction nagelienne est même compatible avec un dualisme cartésien des substances, ou un parallélisme leibnizien. Il suffit que la corrélation entre les propriétés mentales et physiques soit nomique, de sorte à permettre de déduire les descriptions psychologiques à partir des descriptions physiques. Kim le dit clairement :

Si nous supposons que la psychologie a effectivement été réduite en ce sens [c'est-à-dire selon le modèle de Nagel], qu'est-ce que cela montrerait ? La réponse est que cela montrerait bien peu de choses, car selon les théories dualistes classiques concernant la relation corps/esprit – comme par exemple la théorie spinoziste du double aspect, la doctrine de l'harmonie préétablie de Leibniz, le parallélisme ou le monisme neutre – les lois psy-

26. «Cette conséquence est compatible avec une ontologie non physicaliste. [...] Selon cette interprétation, la vérité du réductionnisme *ne* garantit *pas* la généralité de la physique vis-à-vis des sciences particulières» (Fodor, 1974, p. 129, trad. p. 420, italiques de Fodor).

chophysiques seraient disponibles en abondance. En fait, l'existence de telles lois n'est même pas exclue par le dualisme cartésien [...]. Cela signifie que la réductibilité nagelienne du mental est non seulement compatible avec ces théories dualistes, mais qu'elle est en fait impliquée par plusieurs d'entre elles. Si la réduction psychophysique doit être comprise sur la base du modèle de la réduction nagelienne, le réductionnisme n'aurait donc pas de conséquences métaphysiques intéressantes (Kim, 2000, p. 17).

Nous parvenons donc au résultat apparemment paradoxal selon lequel la réduction nagelienne, au lieu d'être une condition trop forte qui prive le mental de son autonomie, s'avère être une condition trop faible, dont la satisfaction ne garantit même pas la vérité du matérialisme. On peut tirer la même conclusion, à plus forte raison, à propos du modèle de la réduction de Schaffner, Hooker et Churchland qui impose des conditions de réductibilité plus faibles.

La raison de cet échec de la réductibilité à garantir la matérialité des propriétés réduites est semblable à la raison pour laquelle la survéance échoue à elle seule à garantir le matérialisme : tout comme la corrélation qui caractérise la survéance, la seule exigence d'une dépendance nomique entre propriétés physiques et mentales ne suffit pas à expliquer leur corrélation. Cela est particulièrement clair lorsqu'on songe au fait que, si l'on conçoit les propriétés mentales comme émergentes au sens de l'émergentisme britannique, la condition nagelienne de connectabilité est satisfaite : selon Broad (1925), les propriétés émergentes sont nomiquement déterminées par les propriétés de base à partir desquelles elles émergent, mais cette dépendance nomique est absolument primitive et susceptible d'aucune explication. Le fait que telle propriété mentale émerge à partir de telle propriété du cerveau est un « fait brut » qu'il faut accepter, pour utiliser une expression d'Alexander (1920), avec « piété naturelle » sans essayer de le comprendre<sup>27</sup>. C'est parce que le modèle de Nagel n'impose pas que les principes de liaison soient à leur

27. Depuis Locke, l'association dans la sensation de certaines qualités secondes avec certaines stimulations sensorielles a été considérée comme le paradigme du caractère mystérieux des propriétés émergentes. Locke exprime l'idée qu'il s'agit d'un « fait nomique brut » imperméable à toute investigation possible, en disant que c'est la volonté impénétrable de Dieu qui apparaît de notre point de vue comme arbitraire en associant tel stimulus à la sensation de l'odeur de la violette, et tel autre à une sensation douloureuse. « Supposons maintenant [...] qu'une violette, par la poussée de ces particules insensibles de matière, de figure et de masse spécifiques, avec des mouvements de degrés et de types différents, produise dans l'esprit les idées de couleur bleue et d'odeur sucrée; car, il n'est pas plus impossible de concevoir que Dieu annexe telle idée à tel mouvement qui n'a aucune similitude avec elle, que de concevoir qu'Il annexe l'idée de *douleur* au mouvement d'un morceau d'acier coupant, qui n'a aucune ressemblance avec cette idée » (Locke, *Essay II*, 8, p. 13, trad. fr. p. 222, italiques dans le texte).



tour explicables et compréhensibles de manière scientifique qu'il laisse la porte ouverte à des conceptions dualistes du mental. En ce sens, le modèle nagelien ne peut satisfaire aux exigences matérialistes : « La réduction, dit Kim, doit rendre intelligible comment certains phénomènes apparaissent à partir de phénomènes plus fondamentaux », mais dans la mesure où les « principes de liaison sont considérés comme des prémisses auxiliaires non expliquées, [la réduction nagelienne] ne fera pas avancer notre compréhension du mental d'un centimètre » (Kim, 1998, p. 96).

#### 4 – Le physique détermine le mental (au sens logique)

À partir du constat que l'existence de lois-pont nageliennes est une exigence trop faible pour garantir le matérialisme, certains philosophes ont tiré la conclusion que le rapport entre corps et esprit doit être conçu comme plus étroit que la seule corrélation nominale<sup>28</sup> : il s'agit d'une relation conceptuelle, accessible *a priori*.

Selon l'hypothèse de Stephen Yablo, les propriétés mentales et les propriétés physiques sous-jacentes sont dans un rapport de *détermination*. Les propriétés mentales sont des déterminables, au sens logique, des propriétés physiques qui sont leurs déterminés. « Les relations mental-physique sont une espèce de relations déterminable-déterminé » (Yablo, 1992, p. 256), en analogie avec la relation de l'écarlate au rouge. L'écarlate est avec d'autres teintes de rouge, comme le carmin et le bordeaux, une propriété déterminée relativement au rouge qui en est la propriété déterminable. D'une manière analogue, suggère Yablo, la propriété qui caractérise un corps humain sur le plan microscopique est une propriété déterminée par rapport à la propriété déterminable qui caractérise l'état de son esprit. Le fait de concevoir le rapport du mental au physique de cette manière est en particulier censé permettre de résoudre le problème de l'exclusion causale.

28. Je laisse de côté deux approches. Selon Causey (1977), seule l'interprétation des principes de liaison comme énoncés d'identité leur donne un réel pouvoir explicatif. Mais nous avons vu que la réalisabilité multiple au sens de la détermination micro-macro est incompatible avec la forme biconditionnelle des principes de liaison, elle-même requise pour l'identité. Churchland (1981) et Bickle (2003) argumentent au contraire que la découverte des mécanismes neurophysiologiques déterminant notre comportement montre l'inadéquation des concepts psychologiques, et conduit donc à l'élimination des propriétés mentales de notre ontologie scientifique qui devient ainsi purement physique. Cette thèse est incompatible avec l'existence même de la psychologie scientifique : l'existence de corrélations nominales parmi les propriétés psychologiques réfute l'éliminativisme.

Voici le problème : dans la mesure où l'on accepte le principe de la clôture causale du physique, chaque événement physique possède une cause suffisante entièrement physique (à chaque instant qui le précède). Or, si tout événement a une cause physique complète, il est difficile de voir quel rôle causal il reste à jouer pour le mental. Il est certainement peu plausible d'envisager que toutes les causes mentales sont des cas de « surdétermination », au sens où tous les événements causés par l'esprit d'un agent ont toujours deux causes complètes, chacune suffisante, à savoir une mentale et une physique. Il n'y a aucune raison indépendante (sauf celle de sauver à tout prix la responsabilité causale du mental) de penser que la causalité mentale relève d'une telle surdétermination systématique alors qu'on ne la trouve nulle part ailleurs dans la nature.

L'hypothèse selon laquelle le rapport entre une propriété mentale et la propriété physique sous-jacente est celui d'une propriété déterminable à une propriété déterminée permet de résoudre ce problème. Supposons que le fait qu'une surface est rouge soit causalement responsable d'une certaine réaction d'un pigeon conditionné au rouge. La responsabilité du fait que la surface soit rouge n'entre pas en compétition causale avec le fait que la surface soit écarlate. Les deux propriétés ont une responsabilité causale pour le même effet. Cela n'empêche pas que l'une d'entre elle mérite davantage d'être considérée comme la cause parce qu'elle est « proportionnelle » à l'effet<sup>29</sup>. Si le pigeon est conditionné au rouge en tant que tel et non pas à une teinte particulière de rouge, la couleur écarlate d'un stimulus est suffisant mais non nécessaire pour déclencher la réponse. La couleur rouge est la cause qui est « proportionnelle » à cette réponse, dans la mesure où la perception de rouge est nécessaire et suffisante pour la réponse. De manière analogue, propose Yablo, lorsque le fait qu'une personne possède une propriété mentale est responsable du fait que cette personne exécute un mouvement corporel, cela n'empêche pas la propriété cérébrale sous-jacente d'être elle aussi causalement responsable du même effet. Dans la mesure où la propriété mentale est en outre proportionnelle à l'effet alors que la seconde est surchargée de détails sans pertinence<sup>30</sup>, il est correct de dire que la propriété mentale est la cause du mouvement corporel.

29. Cf. Yablo (1997).

30. « Les états physiques » dit Yablo, qui réalisent un état mental donné apparaissent, « à la lumière de la proportionnalité [...] comme surchargés de détails microstructurels non nécessaires. (Si ma douleur avait été implémentée d'une manière microphysique différente, l'effet aurait selon toute probabilité néanmoins eu lieu. ) » (Yablo, 1997, p. 254).

Cette conception a également un net avantage par rapport aux relations de survenance et de dépendance nomique en ce qui concerne le matérialisme : les propriétés physiques et mentales relèvent du même genre de propriétés, comme l'écarlate et le rouge. Par conséquent, dans la mesure où les propriétés cérébrales sont matérielles, les propriétés mentales le sont aussi. Cependant, ce qui fait la force de la conception de Yablo, l'étroitesse du rapport physique-mental conçu comme un rapport de dépendance logique, fait aussi sa faiblesse.

Le concept de détermination de Yablo est de nature logique. Yablo précise qu'il s'agit d'une nécessitation par «inclusion» : lorsque Y est dite être déterminée de X qui est sa déterminable, cela ne signifie rien d'autre que le fait que «Y nécessite X (non pas parce qu'il possède une manière métaphysiquement infaillible de le produire mais) parce que X est immanente ou incluse dans Y» (Yablo, 1997, note 22).

Comment peut-on interpréter la notion d'inclusion entre propriétés? Le concept d'inclusion, qui vient de la théorie des ensembles, semble adéquat pour caractériser certains rapports entre concepts ou prédicats : en un sens, une conjonction «inclut» chacun de ses termes. On peut aussi attribuer un sens à l'expression selon laquelle chacun de ses termes est «immanent» à une conjonction ; et lorsque l'on attribue un prédicat conjonctif à un individu, cela implique, ou «nécessite», logiquement les propositions lui attribuant chacun des termes de la conjonction : si  $a$  est un triangle rouge, alors  $a$  est un triangle ; et si  $a$  est un triangle rouge, alors  $a$  est rouge.  $(T\&R) a \rightarrow Ta$  et  $(T\&R) a \rightarrow Ra$ . Ces concepts d'inclusion, d'immanence et de nécessitation, peuvent aussi s'appliquer aux propriétés exprimées par les concepts et dénotées par les prédicats, à condition qu'il s'agisse de propriétés qui ont une «structure conjonctive» : dans la mesure où il y a un sens auquel la propriété déterminée est «composée» de propriétés plus simples, on peut dire de la propriété d'être un triangle rouge qu'elle est «composée» de la propriété d'être un triangle et de celle d'être rouge.

Comment pouvons-nous évaluer la plausibilité de l'hypothèse selon laquelle les propriétés cérébrales possèdent une telle structure conjonctive où l'un des termes correspond à une propriété mentale? Considérons le critère épistémique suivant : si l'on connaît la conjonction, on peut connaître *a priori* les termes de la conjonction. Ainsi je peux tirer, par un effort d'abstraction *a priori*, de ma connaissance de la propriété

conjonctive d'être un triangle rouge, la connaissance des propriétés d'être un triangle et d'être rouge<sup>31</sup>.

Or, selon ce critère, les propriétés mentales ne peuvent pas être des déterminables des propriétés cérébrales. Si la propriété cérébrale *C* avait une structure conjonctive *M&X* où *M* est une propriété mentale, je pourrais obtenir *M* à partir de la connaissance de *C* en faisant abstraction par un effort *a priori*, du «détail superflu» *X*.

Dans l'état actuel de nos connaissances, la relation entre une propriété physique du cerveau et la propriété mentale à laquelle elle donne lieu n'apparaît certainement pas comme purement conceptuelle, au sens où la connaissance de la dernière n'est pas accessible de manière *a priori* à partir de la connaissance de la première. Je peux connaître toutes les propriétés du cerveau d'une personne sur le plan neurophysiologique, sans pouvoir en tirer, par un effort *a priori* d'abstraction, la connaissance de ses propriétés mentales, comme je peux déduire *a priori* du fait qu'un objet est écarlate, le fait plus abstrait que l'objet est rouge.

Un autre critère pour évaluer la proposition de Yablo est d'effectuer des raisonnements contrefactuels : si la relation entre le cérébral et le mental est une relation interne logique, comme c'est le cas si l'hypothèse de Yablo est correcte, il est impossible de *concevoir* qu'un objet possède l'une sans l'autre. Si *M* est une déterminable de la déterminée *C*, alors l'existence de «zombies»<sup>32</sup> est logiquement impossible ou *inconcevable* : un zombie (au sens philosophique) est un objet qui partage toutes les propriétés physiques avec une personne, sans avoir de propriétés mentales. (Souvent, on considère qu'il ne manque au zombie que les expériences qualitatives, les «qualia», ce qui lui permet d'avoir des représentations.) Or, il apparaît clair que l'on peut *concevoir* des zombies<sup>33</sup>.

La concevabilité des zombies suggère qu'il n'y a pas en réalité de lien conceptuel entre les propriétés cérébrales et mentales, et que la relation entre elles n'est pas connaissable *a priori*. D'où l'hypothèse selon laquelle il s'agit d'une relation de «détermination nomique», c'est-à-dire de détermination en vertu de lois de la nature : si *M* est déterminé *par des lois*,

31. Techniquement, la relation entre la propriété conjonctive et chacune de ses composantes est une relation interne : la propriété complexe ne peut pas exister sans ses composantes.

32. Le concept de zombie a été introduit dans la littérature philosophique par Kirk (1974). Cf. Chalmers (2003).

33. En disant cela, on ne préjuge pas la réponse à la question de savoir si de tels zombies soient *réellement* (ou «*métaphysiquement*») *possibles*, c'est-à-dire compatibles avec les lois de la nature.

et non sur la base de la seule logique, à partir des propriétés microscopiques des parties de l'organisme, un zombie est concevable. Il suffit de concevoir un monde possible dans lequel les lois de la nature ne sont pas les lois actuelles. Un tel monde peut bien être métaphysiquement impossible, mais dans la mesure où les lois n'exercent pas des contraintes sur la cohérence de nos concepts, il est concevable. Dans un monde concevable où les lois psychophysiques ne sont pas les nôtres, il y a par exemple des «zombies» qui possèdent des états neuronaux qui *nous* font sentir une douleur, sans pourtant ressentir la moindre douleur. La concevabilité de tels zombies contredit l'affirmation de Yablo selon laquelle

on ne peut même pas poser la question de savoir comment je me serais comporté si l'état cérébral avait été présent en l'absence de la douleur, car l'état cérébral inclut la douleur (Yablo, 1997, p. 258).

Les propriétés cérébrales sont certainement des propriétés structurées complexes: on peut les concevoir, avec un concept introduit par Kim, comme des propriétés «macroscopiques micro-basées» (expression désormais abrégée par «MMB»). L'individu *a* possède une propriété cérébrale *C*, en vertu du fait que les parties de *a*, notamment ses parties neuronales  $a_1, a_2, \dots$  possèdent des propriétés microscopiques  $P_{11}, P_{12}, \dots, P_{n1}, \dots$ . Ces propriétés microscopiques peuvent être des états d'excitation des différents neurones dans le cerveau et des forces de liaison synaptique entre ces neurones. Il est certainement correct que les propriétés cérébrales *C* «déterminent» en un sens les propriétés mentales de l'individu. Cependant, la conception de ces propriétés *C* comme des propriétés structurées n'offre aucune prise à l'application du modèle de Yablo: les termes de la conjonction qui constitue la microbase de *C* n'appartiennent pas au même objet que *C*, mais à ses parties. Par conséquent, en supprimant des termes, on obtient des propriétés qui n'appartiennent pas au même objet que la conjonction; le rapport entre une propriété MMB et une conjonction partielle contenant une partie mais non l'ensemble des termes qui forment la conjonction de la MMB ne relève pas du genre de rapport qui existe entre une propriété déterminée et une propriété déterminable *du même objet*, ce qui rend cette conception du rapport cérébral-mental incompatible avec le modèle de Yablo.

### 5 – Détermination nomique

Dans une perspective matérialiste, les propriétés mentales et les propriétés cérébrales attribuées au cerveau en tant que tout sont toutes des

propriétés MMB au sens de Kim : ce sont des propriétés appartenant à des objets complexes (les personnes, ou autres systèmes cognitifs), en vertu des propriétés de leurs parties (en particulier les propriétés de neurones cérébraux et de récepteurs sensoriels) et des interactions entre ces parties (en particulier les interactions entre neurones grâce aux connexions synaptiques). Il est également indéniable que les premières déterminent les secondes. Seulement, cette détermination n'est pas logique ou conceptuelle, comme le soutiennent Yablo et le néofonctionnalisme de Kim<sup>34</sup>. Il s'agit plutôt d'une relation de détermination *nomique*. Considérons l'ensemble des parties  $a_1, a_2, \dots$  de l'individu  $a$ , avec leurs propriétés  $P_{11}, P_{12}, \dots, P_{n1}, \dots$ . Cet ensemble détermine, *en vertu de la seule logique*, le fait que  $a$  possède la propriété MMB  $C$ , d'être composée de ces parties ayant ces propriétés. Cela suit de la définition que donne Kim d'une macropropriété micro-basée :  $c$ 'est une propriété qui « appartient à un tout en vertu de faits concernant ses parties » (Kim, 1988, p. 142 ; 1993c, p. 124). Dans cette caractérisation, la relation désignée par « en vertu de » est purement logique :

$P$  est une *propriété micro-basée* seulement au cas où  $P$  est la propriété d'avoir des parties propres  $a_1, a_2, \dots, a_n$ , qui sont telles que  $P_1(a_1), P_2(a_2), \dots, P_n(a_n)$ , et  $R(a_1, \dots, a_n)$  (Kim 1997b, p. 292).

Le cas le plus simple d'une propriété microbasée est la propriété d'être « composé de deux parties  $x$  et  $y$  de telle sorte que  $x$  est  $F$  et  $y$  est  $G$  et  $x$  se trouve dans une relation  $R$  par rapport à  $y$  » (Kim, 1988, p. 142 ; 1993c, p. 124).

Mais la seule logique ne garantit pas que la propriété ainsi déterminée soit une propriété réelle au sens où elle entre dans des relations de détermination nomique (autrement dit, dans des lois de la nature)<sup>35</sup>. Cela apparaît clairement lorsqu'on considère un tout méréologique dont les parties n'interagissent pas du tout physiquement, par exemple parce que certaines de ses parties se trouvent sur Terre et certaines se trouvent sur une galaxie lointaine. En vertu de la définition de la MMB, un tel tout « discontinu » possède néanmoins une propriété MMB. Cependant, celle-

34. Je dois laisser ici de côté l'analyse de la conception du rapport entre le mental et le physique récemment proposée par Kim qui réinterprète le fonctionnalisme de manière à le rendre compatible avec la réduction des propriétés mentales aux propriétés physiques.

35. Selon le « critère causal de réalité », une propriété est réelle si et seulement si son instantiation par un événement (ou un objet à un instant donné) change les relations causales ou, plus généralement, nomiques de cet événement. Cf. Kistler (2002).

ci n'est pas réelle : elle ne donne aucun pouvoir causal à son possesseur et n'est nomiquement responsable d'aucune autre propriété.

En général, la propriété MMB C n'a donc pas de réalité au sens nomique et causal : ce n'est pas une propriété «réelle». Si on peut néanmoins l'appeler «propriété», c'est au sens faible où chaque prédicat, même arbitrairement composé notamment par les opérations logiques de disjonction et de conjonction, désigne une «propriété»<sup>36</sup>. La propriété MMB envisagée plus haut appartient à cette dernière catégorie : elle appartient au tout méréologique simplement parce que ses parties microscopiques ont certaines propriétés, sans que cela implique qu'elle figure elle-même dans des lois. En particulier, la conception du tout qui possède la MMB ne dépend pas de l'existence d'une interaction physique quelconque entre ses parties. En revanche, les propriétés mentales sont des propriétés réelles et causalement efficaces, notamment dans l'action. Leur réalité peut recevoir une explication nomologique à partir de lois de constitution : les propriétés mentales ne peuvent appartenir qu'à des objets physiquement intégrés. On peut expliquer leur existence en termes d'une détermination par des lois de la nature. Cependant, elles ne sont pas déterminées *logiquement* à partir de C, mais à partir de l'ensemble des propriétés des parties de leur possesseur, en vertu des lois qui régissent les interactions entre ces parties. Ce sont les lois d'interaction entre les électrons et les protons d'une molécule qui déterminent les propriétés chimiques qui sont des propriétés réelles de la molécule en tant que tout. De la même manière, on peut faire l'hypothèse que ce sont les lois d'interaction entre les parties de l'organisme qui déterminent, nomiquement ou «naturellement», les propriétés cognitives réelles de l'organisme. Ces propriétés sont des propriétés causalement efficaces de premier ordre : la représentation de l'endroit où se trouve la cerise contribue causalement aux mouvements corporels nécessaires pour la manger. Elles sont déterminées par les interactions entre les propriétés d'activation des neurones dans l'aire cérébrale responsable de la carte cognitive de l'organisme<sup>37</sup>.

36. Ce que j'appelle les «propriétés réelles» correspond aux propriétés «éparses» (*sparse*) de la terminologie de Lewis (1983). Lewis appelle «abondantes» les propriétés qui existent grâce à un prédicat quelconque qui les désigne, sans être soumises à l'exigence d'entretenir des relations de dépendance nomique. Cf. aussi Mellor (1991).

37. La nature exacte de ces interactions n'est pas encore connue. Mais il y a des raisons de penser qu'il s'agit d'un processus correspondant à un moyennage. Cf. Gallistel (1990), Churchland et Sejnowski (1992).

## 6 – Conclusion

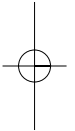
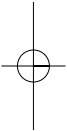
Nous avons suggéré de concevoir les propriétés mentales ou cognitives comme des propriétés réelles au sens où elles ont des pouvoirs causaux propres, distinctes des pouvoirs causaux des propriétés des neurones individuels. Cela n'empêche pas qu'elles soient en principe réductibles : la réduction consiste dans la découverte des lois qui font émerger la propriété macroscopique. Ce modèle est censé couvrir tous les paliers de l'échelle qui va du microscopique au macroscopique, jusqu'à la détermination des propriétés mentales par l'ensemble des propriétés d'activation des neurones et de leurs interactions. Il est censé fournir une interprétation ontologique de la réduction : les réductions scientifiques sont fondées, non simplement sur l'existence de corrélations, mais sur la découverte de relations de détermination nomique des propriétés macroscopiques, à partir de leurs parties. Cela montre, à l'encontre de la thèse de Causey (1977), que l'identification n'est pas la seule manière satisfaisante d'interpréter la possibilité d'une réduction sur le plan ontologique.

Cette conception est matérialiste en concevant toutes les propriétés réelles, y compris mentales, comme étant en principe réductibles. À la différence du modèle nagelien, notre conception exige que les lois en vertu desquelles les propriétés de haut niveau sont déterminées par les propriétés de bas niveau et leurs interactions soient elles-mêmes dérivables à partir des lois d'interaction de bas niveau. C'est cette contrainte qui garantit la matérialité des propriétés de haut niveau. Elle s'oppose à l'éliminativisme dans la mesure où elle reconnaît que les propriétés cognitives sont des propriétés réelles et causalement efficaces. Elle s'oppose à la théorie de l'identité dans la mesure où les propriétés mentales ne sont que déterminées, en vertu de lois de la nature, par les propriétés neurophysiologiques sous-jacentes, sans être identiques avec elles. Elle partage avec le physicalisme la conviction que tous les objets sont constitués uniquement de parties physiques ; cependant, il serait mal approprié de l'appeler « physicaliste » dans la mesure où elle suppose que nombre de phénomènes, notamment d'ordre psychologique, ne peuvent pas être expliqués même en principe dans le cadre de la physique : leur explication doit nécessairement faire appel à des propriétés non-physiques qui possèdent leurs pouvoirs causaux propres tout en étant réductibles.





La relation entre le physique et le mental est, selon cette conception, moins étroite qu'une relation conceptuelle. Cela rend compte du fait clairement établi par Descartes que l'on peut *concevoir* l'existence du corps sans l'existence de l'esprit<sup>38</sup>. Or cela n'empêche qu'il s'agit d'une relation qui a la force modale de la nécessité métaphysique : les mondes où il existe des zombies sont concevables mais ils sont métaphysiquement impossibles dans la mesure où de tels mondes violent les lois psychophysiques qui sont métaphysiquement nécessaires comme toutes les lois de la nature<sup>39/40</sup>.



---

38. Bien entendu, Descartes insiste surtout sur la possibilité inverse de concevoir l'existence de l'esprit sans celle du corps.

39. Cette dernière thèse est bien entendu controversée. La place nous manque ici pour la justifier. J'ai essayé de le faire ailleurs. Cf. Kistler (2002) et Kistler (à paraître, b).

40. Je remercie Luc Faucher, Georges Jobert, Veronica Ponce et Jürgen Schröder pour leurs remarques critiques.

