

## CAUSALITÉ MENTALE ET NIVEAUX DE CAUSALITÉ

Considérons les explications suivantes. C'est parce que le morceau de charbon était rouge qu'il m'a brûlé. Mais ce n'est pas vrai. Le morceau de charbon m'a brûlé parce qu'il était à une certaine température. C'est parce que les parents avaient les yeux bleus que l'enfant a les yeux bleus. Mais ce n'est pas vrai. L'enfant a les yeux bleus parce que les parents étaient porteurs de certains gènes. La couleur du charbon, celle des yeux bleus sont des pseudo-processus. Les causes et les généralisations invoquées sont de pseudo-causes, qui résultent d'une cause commune véritable. Comparons maintenant ces explications avec celle-ci : c'est parce que l'agent croyait que cette carte était rouge qu'il l'a posée à côté des autres cartes. Nous n'avons pas ordinairement tendance à considérer ces explications du comportement comme reposant sur des pseudo-processus<sup>1</sup>. Pourtant ces causes « mentales » sont peut-être des pseudo-processus. Par exemple Changeux a récemment soutenu, à propos d'une expérience de sélection de cartes comparable à celles que les psychologues donnent à des sujets pour tester leurs compétences à manipuler certaines règles, qu'il y avait des « neurones-règles » et même des « neurones d'erreur » systématiquement associés à leurs comportements durant ces tests<sup>2</sup>. La question que je voudrais poser ici est celle de savoir si l'on peut dire que les causes mentales que

1. Cf. G. Segal et E. Sober, *The Causal Efficacy of Content*, *Philosophical Studies*, 63, 1-30; sur la notion de « pseudo-processus », cf. W. Salmon, *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, 1984.

2. J.-P. Changeux, *Les neurones de la raison*, *La Recherche*, juin 1992.

nous invoquons dans nos explications ordinaires du comportement sont des pseudo-processus. Si elles sont des pseudo-processus, elles ne jouent pas réellement un rôle causal et sont épiphénoménales.

Il y a cependant plusieurs formes d'épiphénoménisme, plusieurs versions du problème. La version qui nous intéresse ici est la suivante. Nous pourrions vouloir dire, au sujet de nos exemples, qu'ils illustrent non pas des cas dans lesquels il y a une pseudo-cause et une cause véritable, mais des cas dans lesquels nous avons affaire à deux niveaux d'explication causale. Après tout, c'est une loi que, quand le morceau de charbon atteint une certaine température, il rougit. De même si nous admettons comme critère des énoncés causaux qu'ils soient « étayés » par des énoncés contrafactuels, alors on peut dire que si le morceau de charbon n'avait pas été rouge, il ne n'aurait pas brûlé. De même nous pouvons dire que, si l'agent n'avait pas cru que la carte était rouge, il ne l'aurait pas groupée avec les cartes de même couleur. Il y a donc apparemment une certaine vérité de ces attributions de causes, même quand nous avons des raisons de penser que ce sont des pseudo-processus, et on peut suggérer qu'il y aurait en ce sens différents niveaux d'explication causale. Ce langage des niveaux est familier. Des énoncés décrivant des propriétés de la température d'un gaz peuvent énoncer des propriétés causales au niveau phénoménologique, mais les propriétés causales du système peuvent être envisagées au niveau moléculaire. Un sociologue peut dire que l'augmentation du chômage est la cause de l'augmentation de la délinquance dans une ville, mais on peut dire aussi que l'explication peut être donnée à un niveau microsociologique, en invoquant un ensemble de facteurs individuels tenant aux situations et aux attitudes psychologiques des agents individuels. De même la distinction entre macro-économie et micro-économie est familière. Enfin, en psychologie cognitive, on distingue couramment entre les explications au niveau intentionnel (reposant sur des attributions de contenu d'attitudes propositionnelles) des explications au niveau subintentionnel, invoquant des représentations et des processus à un niveau « inférieur ». La question traditionnelle que posent ces distinctions de niveau est : Sont-ils autonomes et chaque niveau énonce-t-il la présence d'un facteur causal authentique, ou bien y a-t-il un niveau fondamental d'explication causale en sorte que les explications des niveaux « supérieurs » doivent se réduire aux explications des niveaux inférieurs ? Admettons, pour le moment, le critère suivant de la distinction des niveaux d'explication causale. Supposons que deux sortes de propriétés soient causalement pertinentes pour un certain effet, par

exemple la fragilité d'un verre et sa structure moléculaire. Nous dirons que la fragilité du verre est une propriété pertinente d'ordre supérieur parce qu'elle est elle-même produite par la structure moléculaire du verre, et non pas inversement. Mais la difficulté posée par cette distinction de niveaux est la suivante : si nous supposons que chaque niveau identifie une cause spécifique, n'allons-nous pas nous trouver dans la situation où il y aurait deux causes distinctes concourant à la production d'un même événement (disons, ici, le bris du verre), c'est-à-dire dans un cas de surdétermination ? Et si nous disons que le niveau supérieur d'explication (qui invoque la propriété de fragilité) est en quelque sorte dérivé par rapport au niveau inférieur, et moins fondamental, ce dernier ne va-t-il pas exclure explicativement<sup>1</sup> le niveau supérieur ? Mais dans ce cas, quel sens cela aurait-il de parler de l'autonomie des deux niveaux ? Ne vaut-il pas mieux dire tout simplement que tous les niveaux supérieurs doivent se réduire à un seul, le niveau le plus inférieur, que nous présumons être le niveau microphysique ?

Le problème de l'« exclusion explicative » ne se pose pas si nous adoptons la thèse du physicalisme réductionnisme, entendue au sens suivant : toutes les propriétés causalement pertinentes invoquées à un niveau « supérieur » d'explication, et toutes les régularités et lois causales invoquées à un tel niveau doivent se réduire à des propriétés et des lois du niveau inférieur fondamental décrites par la physique. Mais le réductionnisme physicaliste n'a pas bonne presse dans la philosophie contemporaine. En économie, la loi de Fischer énonce que le niveau des prix dans une économie est fonction de la masse monétaire dans cette économie, à condition que les échanges monétaires aient une certaine rapidité et que la quantité des biens et des services demeure constante. Si nous ignorons cette restriction, nous dirons que, d'après le réductionnisme physicaliste, la loi :

Si un événement satisfait le prédicat « est une augmentation de la masse monétaire », alors il produira causalement un événement qui satisfait le prédicat « est une augmentation des prix ».

Admettons qu'il y ait, comme le veut la conception réductionniste classique<sup>2</sup>, des « lois-ponts » entre les généralisations de ce type et des généralisations au niveau physique. Alors, si l'économie est

1. Selon la terminologie de J. Kim, Explanatory Exclusion and the Problem of Mental Causation, in E. Villanueva (ed.), *Information, Semantics and Epistemology*, Blackwell, 1990, p. 36-56.

2. Celle de E. Nagel, *The Structure of Science*, Routledge, 1961.

réductible à la physique, on devra pouvoir dériver des lois comme celle de Fischer de lois liant des événements instanciant des prédicats physiques<sup>1</sup>. Quelqu'un qui connaîtrait tous les faits physiques pourrait expliquer tout ce que l'économiste explique. Mais même si c'était vrai, il semble que cela n'aurait aucun intérêt. Car il serait extrêmement encombrant de faire appel à des lois physiques pour expliquer les faits économiques. C'est une première raison de s'opposer au réductionnisme. Une seconde raison, couramment alléguée, est que non seulement ce n'est pas intéressant explicativement, mais faux, parce qu'il n'est pas vrai que toute propriété non physique d'un événement sera réalisée invariablement par une seule et même propriété physique. Par exemple Dennett nous invite à considérer des martiens superphysiciens laplaciens « capables de comprendre l'activité de Wall Street au niveau microphysique » :

« Prenez le cas où les martiens observent un agent de change, décidant de donner un ordre d'achat de 500 parts de General Motors. Ils prédisent les mouvements exacts de ses doigts quand il compose le numéro de téléphone, et les vibrations exactes de ses cordes vocales quand il émet son ordre. Mais si les Martiens ne voient pas qu'on aurait pu substituer aux objets particuliers en présence un nombre indéfini de structures différentes de mouvements de doigts et de vibrations de cordes vocales — même les mouvements d'un nombre indéfini d'individus différents — sans perturber les événements extérieurs du marché boursier, alors c'est qu'ils ont été incapables de déceler une structure véritable du monde qu'ils sont en train d'observer. Tout comme il y a une infinité de manières d'être une bougie — et on n'a pas compris en quoi consiste un moteur à explosion tant qu'on ne réalise pas que l'on peut visser sur ces douilles toute une variété de dispositifs différents sans affecter la performance du moteur —, il y a de même une infinité de manières de donner des ordres d'achat de la General Motors, et il y a des variétés de dispositifs propres à chaque société qui sont autant de douilles dans lesquelles l'une de ces diverses manières produira à peu près le même effet que n'importe quelle autre. Il y a aussi des pivots spécifiques à chaque société, pour ainsi dire, où la direction que prennent les gens dépend du fait qu'ils croient que *p*, ou désirent que *q*, et ne dépend d'aucune façon de l'infinité des autres manières par lesquelles ils peuvent être semblables ou différents. »<sup>2</sup>

Dennett avance ici un point familier : il peut y avoir une grande variété de réalisations physiques différentes d'une même propriété de niveau supérieur, qu'il s'agisse de la propriété d'ordonner l'achat

1. J'emploie ici la notion d'instanciation en un sens à la fois proche du sens usuel d'exemplification et de réalisation d'une propriété dans un particulier (ou dans une autre propriété), et du sens logique (où l'on dit, par exemple qu'un nom propre instancie une variable d'individu).

2. D. Dennett, *The Intentional Stance*, MIT Press, 1987, chap. 3, trad. franç. P. Engel, *La stratégie de l'intention*, p. 40-41.

de 500 actions ou de la propriété d'être une bougie ou la propriété de croire que *p*. C'est ce que l'on appelle la « réalisabilité multiple » de ces propriétés<sup>1</sup>.

Le phénomène de la réalisabilité multiple paraît être une raison décisive pour maintenir une distinction entre niveaux d'explication. Mais il n'interdit pas de maintenir, sur le plan ontologique, le physicalisme. On peut très bien soutenir que, même si les propriétés de niveau supérieur, celles qui sont décrites par telle ou telle « science spéciale », telle que l'économie, la sociologie, la biologie ou la psychologie, ne sont pas réductibles à des propriétés physiques fondamentales, il ne s'ensuit pas que l'on ne puisse pas dire que les événements qui instancient ces différentes propriétés sont des événements physiques, et il ne s'ensuit pas que l'on ne puisse pas dire que les propriétés de niveau supérieur non physique ne dépendent pas, d'une certaine manière, des propriétés physiques. On peut soutenir en effet la thèse suivante :

*Survenance* : Il est nécessairement vrai que si deux mondes sont exactement similaires sous tous leurs aspects physiques, alors ils sont exactement similaires sous tous leurs autres aspects<sup>2</sup>.

En d'autres termes, il ne peut pas y avoir de différence dans les aspects non physiques de deux mondes sans qu'il y ait une différence dans leurs états physiques, ou : tout ce que Dieu a à faire pour fixer les faits non physiques relatifs à un monde consiste à fixer les faits physiques relatifs à ce monde. Or la survenance n'implique pas le réductionnisme. Si les faits économiques surviennent sur des faits physiques, alors chaque fois qu'il y a une augmentation de la masse monétaire, il y aura un ensemble de conditions physiques suffisantes pour produire cette augmentation. Mais ces conditions physiques ne seront pas obligatoirement les conditions nécessaires pour qu'il y ait augmentation de la masse monétaire. La survenance est une relation asymétrique. Si ma douleur est surve-

1. Cf. Fodor, *Special Sciences, Synthese*, 1974, trad. franç. in P. Jacob (éd.), *De Vienne à Cambridge*, Paris, Gallimard, 1980.

2. Sur la relation de survenance, cf. Davidson, *Mental Events*, in *Essays on Actions and Events*, Oxford, 1980, trad. franç. P. Engel, *Actions et événements*, et ma note à cette traduction p. 286-287. Cf. également D. Laurier, *L'anomalisme du mental et la dépendance psychophysique*, in D. Laurier et F. LePage (éd.), *Langage et intentionnalité*, Montréal, Paris, Vrin, Bellarmine, 1992; P. Engel, *États d'esprit*, Aix-en-Provence, Alinéa, 1992, 2<sup>e</sup> éd. augmentée, *Introduction à la philosophie de l'esprit*, Paris, La Découverte, 1994. J'emploierai la terminologie suivante : je dirai qu'une propriété (ou un ensemble de propriétés) *P* survient sur une propriété *Q*, que la propriété *Q* subvient sous, ou est subvenante sous la propriété *P*.

nante sur des états neuronaux, quand Dieu met mon esprit dans certains états neuronaux, il est certain de m'infliger de la douleur. Mais si Dieu m'inflige de la douleur, il n'y a aucune garantie qu'il me mettra dans ces états neuronaux, car il peut y avoir de nombreux états physiques qui chacun constituent la douleur.

La survenance est une relation ontologique, et pas une relation épistémologique. Elle ne nous dit rien du pouvoir explicatif de la physique. Elle ne mentionne même pas la notion de cause. Elle implique cependant quelque chose quant à la causalité physique. Supposons que le fait que j'ai une certaine douleur (disons une migraine) soit la cause du fait que je me prends la tête entre les mains. Selon la survenance, j'ai ma migraine en vertu du fait que je suis dans un certain état physique, disons un état neuronal. Etant donné que le fait que je suis dans cet état neuronal est le fait que j'ai cette douleur, alors le fait que je suis dans cet état neuronal est la cause de ce mouvement. Nous pouvons généraliser, et soutenir ceci :

*Prégnance causale* : Si un événement non physique S est la cause d'un autre événement non physique S', alors il y a des états physiques P et P' subvenants tels que P est la cause de l'occurrence de P'.

Mais cette thèse de prégnance causale n'implique rien quant à la manière dont P explique causalement P'. Nous devons distinguer la relation causale, qui est une relation entre événements, et l'explication causale, qui est une relation entre descriptions d'événements. Le fait que A cause B est une chose, et le fait qu'une description d de A explique causalement l'occurrence de B sous une description d' est une autre chose. Par exemple si l'ouragan Donald est la cause de la destruction de la Basse-Terre, et si l'ouragan Donald est l'événement mentionné à la page 14 du *Monde* du 2 décembre, alors que la destruction de la Basse-Terre est l'événement décrit à la page 20 du *Monde* du 3 décembre, il ne s'ensuit pas que l'événement mentionné à la page 14 du *Monde* du 2 décembre explique causalement l'événement mentionné à la page 20 du *Monde* du 3 décembre<sup>2</sup>. Le fait que les événements soient décrits ainsi n'a aucune pertinence explicative pour rendre compte de leur lien causal. En d'autres termes, la relation causale entre événements est extensionnelle, alors que la relation d'explication causale est intensionnelle. Cette distinction nous permet d'ajouter une thèse à la thèse de survenance : qu'un seul et même événement peut être

décrit sous diverses descriptions et entrer en relation causale avec un autre événement. Dans notre exemple, l'événement qu'est l'ouragan peut être décrit comme l'événement mentionné à la page 14 du *Monde* du 2 décembre, et comme l'événement mentionné à la page 20 du *Monde* du 3 décembre. Il n'est pas moins le même événement, et cause, comme tel, un autre événement qui est la destruction de la Basse-Terre. Dans le cas des événements mentaux, nous pouvons dire qu'un seul et même événement, décrit à la fois comme *ma douleur* et comme un certain état neuronal, est la cause d'un autre événement, le fait que je me preme la tête entre les mains.

Même si la survenance des propriétés non physiques sur les propriétés physiques n'entraîne rien quant à l'explication causale des unes par les autres, nous pouvons vouloir dire ceci : chaque fois qu'il existe une explication causale faisant appel à des propriétés non physiques de niveau supérieur, il existe une explication causale des événements instanciés par les propriétés supérieures faisant appel à des propriétés de niveau inférieur instanciées par ces événements. Mais il n'y a aucune raison, à partir de là, de supposer que l'on pourra réduire les explications de niveau supérieur aux explications de niveau inférieur, et établir des corrélations systématiques entre propriétés de niveau supérieur et des propriétés de niveau inférieur. La raison pour laquelle il n'y a aucune raison de supposer cela est à nouveau la réalisabilité multiple.

Appelons cette thèse celle du *physicalisme non réductionniste*. C'est un physicalisme ontologique en vertu de la thèse de survenance. Mais ce n'est pas un réductionnisme en vertu de la non-réductibilité des explications de niveau supérieur aux explications de niveau inférieur. Cette conception préserve donc l'idée de l'autonomie des niveaux explicatifs.

Il y a en fait deux versions distinctes du physicalisme non réductionniste, selon le statut que l'on donne aux explications de niveau supérieur. On peut en effet soutenir, selon une conception classique de l'explication causale, que ces explications sont des explications causales parce qu'elles reposent sur l'existence de lois ou de régularités nomologiques entre propriétés de niveau supérieur. Les défenseurs de cette thèse soutiennent cependant que ces lois ne sont pas des lois « strictes » ou exactes, mais des lois approximatives ou *ceteris paribus*, qui ne sont vraies que relativement à certains paramètres<sup>3</sup>. Selon une autre version du physicalisme

1. Je m'inspire ici de D. Owens, *Levels of Explanation*, *Mind*, 1989.

2. Cf. Davidson, *Actions, reasons and causes*, in Davidson, *op. cit.*

3. Cf. par exemple Fodor, *op. cit.*

lisme non réductionniste, les explications causales de niveau supérieur, et en particulier les explications causales faisant appel à un vocabulaire mental, ne sont pas des explications nomologiques<sup>1</sup>. Cette distinction entre ces deux sortes de conceptions est importante. Mais le problème que je veux poser à présent est indépendant de cette distinction.

Le problème est le suivant. Bien que le physicalisme non réductionniste préserve l'autonomie des niveaux d'explication non physiques par rapport au niveau d'explication physique, la question peut se poser de savoir si les propriétés mentales invoquées dans les explications mentalistes usuelles ont bien des pouvoirs causaux indépendants des pouvoirs causaux des propriétés physiques. La réponse à cette question est en un sens triviale, si l'on admet le principe de survenance et celui de prégnance causale. Les pouvoirs causaux d'une propriété mentale *dépendent* des pouvoirs causaux des propriétés physiques sur lesquelles elles surviennent. En ce sens, les propriétés mentales n'ont pas d'autonomie ontologique. Mais il est à craindre aussi qu'elles n'aient pas d'autonomie explicative. Nous supposons que certaines propriétés microphysiques sont causalement pertinentes pour la production d'un certain événement décrit en termes de propriétés supérieures. Cet événement peut être un changement de couleur ou de forme, il peut être le mouvement d'un objet, il peut être une action. Nous identifions certaines propriétés non microphysiques de cet événement comme causalement pertinentes : la malléabilité d'une gomme qui produit sa flexion, le coup de pied envoyé dans la balle qui la projette en avant, la croyance et le désir d'un agent qui produisent son action. Mais comment ces propriétés peuvent-elles être causalement pertinentes si nous disons aussi que les propriétés microphysiques sont causalement pertinentes ? Comment peuvent-elles collaborer de concert pour produire l'événement en question ? Si nous admettons le principe que j'ai appelé le principe de l'exclusion explicative, cela n'a pas de sens de dire que deux niveaux explicatifs rivaux pourraient collaborer : ils doivent s'exclure mutuellement. Et si nous admettons l'asymétrie de la relation de survenance, le niveau microphysique doit être le plus fondamental : il doit identifier les vraies causes. Nous revenons ici à l'idée que les descriptions des propriétés supposées causalement pertinentes au niveau supérieur doivent être

des pseudo-explications et qu'elles identifient des événements impuissants causalement ou purement épiphénoménaux.

Prenons pour exemple ce morceau de cire qui vient d'être tiré de la ruche...<sup>1</sup> Le morceau de cire est placé sur une grille métallique et, sous l'influence de la chaleur, il change de forme et commence à tomber par un trou dans la grille. Le changement de forme, disons-nous spontanément, est la cause de sa chute à travers la grille. Or on peut supposer que dans ce cas c'est le changement dans l'arrangement spatial des parties microphysiques de la cire qui est la cause correspondante au niveau microphysique. Quelle est donc la relation entre le changement de forme et le changement dans la structure des parties physiques ? La réponse la plus simple, qui découle de la conception physicaliste non réductionniste avancée à l'instinct, consiste à dire que nous avons affaire ici à un seul et unique événement, décrit à un niveau comme un changement de forme, et à un autre niveau comme un changement dans la disposition des parties du morceau de cire. Nous disons aussi que les propriétés identifiées à un niveau surviennent sur les propriétés décrites à l'autre niveau. Mais supposons qu'au moment où changent les parties du morceau de cire le morceau change aussi de volume. Mais si l'on dit que le changement dans la disposition des parties est le même événement que le changement de forme du morceau de cire, alors nous devrions dire aussi que c'est le même événement que le changement de volume. Si nous disons aussi que le changement de forme survient sur le changement de la disposition des parties, nous devrions également dire qu'il survient sur le changement de volume. Mais cela nous conduit à dire que non seulement le changement de forme est causalement pertinent pour la chute de la cire à travers la grille, mais aussi que le changement de volume est causalement pertinent pour cet effet. Mais ce n'est certainement pas ce que nous voulons dire, car le changement de volume n'a aucune pertinence explicative pour expliquer le passage de la cire à travers la grille. Nous pouvons même vouloir dire qu'elle passe à travers la grille *en dépit du fait* qu'elle augmente en volume. Le partisan de la thèse de survenance peut cependant répondre ceci. Le fait que l'on décrive un certain événement tantôt sous la description *changement de forme* et tantôt sous la description *changement de volume* n'implique pas que cet événement, en tant que tel, et pour ainsi dire indépen-

1. En particulier Davidson, *op. cit.*, dont j'ai examiné les difficultés dans *Actions, raisons et causes mentales, Revue de philosophie et de théologie*, 1992, 3, vol. 124, 305-323.

1. Cf. P. Petit, *Pertinence causale et identité des événements, Raisons pratiques*, 2, 1991, p. 57-73, 1992.

damment de ses descriptions, n'est pas la cause de l'événement décrit sous la description *chaue de la cire à travers la grille*. Cette réponse est correcte, étant donné la distinction entre relation causale et explication causale. Mais elle montre que la thèse de survenance, et la thèse de l'identité des événements décrits à divers niveaux, ne nous dit rien de la manière dont nous sélectionnons les propriétés causalement pertinentes des événements en questions, c'est-à-dire rien de la manière dont nous pouvons expliquer pourquoi il y a une telle relation causale. Or notre problème est précisément celui de savoir comment nous pouvons relier l'un à l'autre, dissocier, ou assimiler les deux (ou plusieurs) niveaux d'explication.

Il existe une autre réponse possible à notre problème. Nous pouvons soutenir qu'il nous suffit d'identifier les pouvoirs causaux des propriétés que nous décrivons à un niveau supérieur, sans nous préoccuper de la nature exacte de leurs réalisations physiques à un niveau inférieur. Après tout, n'est-ce pas exactement ce que nous faisons quand nous employons des termes dispositionnels tels que « malléable », « fusible », ou « solide », ou quand nous employons des termes fonctionnels tels que « être une aile » ou « être une serrure » ? Quand nous identifions de telles propriétés comme ayant un pouvoir causal, ce que nous voulons dire, c'est qu'elles occupent un certain rôle causal typique, qui peut être réalisé par diverses structures physiques sous-jacentes. Une chaîne d'antivol pour une bicyclette ou une motocyclette peut être dans deux états, ouverte ou fermée<sup>1</sup>. Nous savons que, quand nous alignons les deux embouts de la chaîne d'antivol, elle s'ouvre. Mais nous pouvons apprendre qu'en fait la chaîne est faite d'une rangée de disques superposés, et que l'alignement des embouts a comme effet typique de provoquer l'alignement des disques, qui est ainsi responsable causalement de l'ouverture de la chaîne. L'alignement des disques est typiquement la cause de l'ouverture de la chaîne. Pourquoi ne pas en dire autant de propriétés telles que les croyances et les désirs des agents ? On peut les définir par les rôles fonctionnels qu'ils occupent, c'est-à-dire par leurs pouvoirs causaux, sans se préoccuper de la manière exacte dont ils sont réalisés physiquement. On sait que c'est là ce que proposent les fonctionnalistes contemporains en philosophie de l'esprit. On préserve ainsi le physicalisme, la sur-

venance, et l'autonomie des explications causales fonctionnelles par rapport aux réalisations physiques.

Mais cela ne résout pas notre problème. D'une part parce que dans l'exemple de la chaîne d'antivol c'est bien la présence des petits disques superposés qui est la cause de l'état d'ouverture ou de fermeture de la chaîne. Dire que cet état fonctionnel a un rôle causal ne nous dit pas en quoi il a un rôle causal, et l'état peut en ce sens être aussi bien inerte causalement et épiphénoménal. Et d'autre part, on ne voit pas en quoi le fait de spécifier l'état pour-rait expliquer toutes les manifestations causales de l'état en question. Supposons qu'une chaîne d'antivol normale ait six disques alignés et qu'elle s'ouvre dans ces circonstances, mais qu'on ait affaire, dans une circonstance spéciale, à une chaîne qui a seulement cinq disques alignés, mais que cela ne suffise pas pour qu'elle soit ouverte, parce qu'il faut une condition supplémentaire (par exemple qu'on prononce un mot magique : « Chaîne ouvre-toi »). Dans ce cas allons-nous dire que la chaîne est dans l'état d'être ouverte si ses disques sont alignés, ou bien qu'elle a un dysfonctionnement quelconque ? Savoir si elle a un dysfonctionnement quelconque, c'est savoir quel est l'état physique spécifique qui le provoque. La même chose est vraie des états mentaux. Si tel et tel désir et telle croyance ont telle ou telle manifestation typique, conduisant l'agent à agir de telle ou telle manière, mais que l'agent est paralysé, l'agent a-t-il ou non le désir et la croyance, est-il ou non disposé à agir ? La simple théorie de l'identité du rôle causal avec un état physique qui le réalise ne suffit pas. En d'autres termes, si un état a de nombreuses réalisations, ce sont ces réalisations et non pas l'état lui-même qui sont causalement efficaces. Nous avons donc de bonnes raisons de penser que la connaissance précise de la condition physique sous-jacente est nécessaire. Et c'est ce que répondent en général au fonctionnaliste les neurophysiologistes qui étudient les lésions cérébrales.

Quelle solution peut-il rester au physicaliste non réductionniste s'il veut continuer à la fois à maintenir la thèse de survenance et l'autonomie des niveaux explicatifs, c'est-à-dire s'il ne veut pas retomber dans le réductionnisme ? On a proposé la solution suivante<sup>1</sup>. Supposons que nous ayons affaire à un cas où nous n'avons pas de doute sur la pertinence causale d'une propriété — ou d'un

1. Exemple emprunté à D. Lewis, *An Argument for the Identity theory*, in *Philosophical Papers*, Oxford University Press, vol. 1, 1983.

1. F. Jackson et Pettit, *Functionalism and Broad Content*, *Mind*, 1988, Program Explanation, a General Perspective, *Analysis*, 1990 ; Causation in the Philosophy of Mind, *Philosophy and Phenomenological Research*, 1990.

groupe de propriétés — décrites à un certain niveau N pour l'occurrence d'un certain effet E. Supposons que nous nous demandions comment une propriété P, à un niveau supérieur, peut être simultanément pertinente causalement pour E. Pour reprendre l'exemple du morceau de cire, comment si c'est le changement des parties microphysiques de la cire qui cause sa chute à travers la grille ce changement peut-il aller de pair avec la capacité qu'à le morceau de changer de forme, son caractère fusible ? Alors on peut proposer le modèle explicatif suivant. La propriété P sera causalement pertinente si et seulement si :

- 1 / L'instanciation de P implique non causalement l'instanciation de certaines propriétés — peu importe exactement lesquelles — au niveau inférieur N, au sens où l'instanciation des propriétés N « réalisent » P ;
- 2 / Les propriétés N associées aux instantiations de P sont en général causalement pertinentes pour l'occurrence de E ;
- 3 / Les propriétés N associées à l'instanciation effective de P sont effectivement causalement pertinentes pour l'occurrence de E.

Que veut dire la mention du fait que l'instanciation de P implique « non causalement » l'instanciation de certaines propriétés ? Elle veut dire qu'elle n'est pas la cause de l'occurrence des événements instantiant ces propriétés, mais qu'elle les rend seulement probables. La malléabilité d'une gomme n'est pas la cause de l'occurrence de telles structures microphysiques qui expliquent causalement le fait qu'elle se torde, mais elle la programme, au sens où elle assure qu'il y aura, au niveau inférieur, des événements de ce type. Nous pouvons dire alors que l'instanciation des propriétés P garantit la *pertinence causale* de ces propriétés, mais pas leur *efficacité causale* ou le *processus causal* mis en œuvre. En d'autres termes, les explications de niveau supérieur ne citent pas des causes effectives, mais elles *pointent vers*, ou désignent, sans les nommer, des explications de niveau inférieur qui mentionnent des causes effectives. Cela nous permet de résoudre la difficulté soulevée plus haut. Dans le cas du morceau de cire, on dira que le changement de forme programme le fait que la cire tombera à travers la grille, alors que le changement de volume ne le fait pas.

C'est une solution intéressante, parce qu'elle nous permet de voir comment deux explications causales distinctes peuvent collaborer dans l'explication d'un même effet, sans qu'il y ait exclusion explicative ni surdétermination. Appliqué à la question de la causalité mentale, ce modèle explicatif conduit à dire que les états que postulent les explications psychologiques ordinaires du comportement ont précisément ce rôle programmeur d'événements neuro-

physiologiques ou situés à un niveau inférieur aux propriétés mentales mentionnées.

Pourtant je ne pense pas que ce modèle résolve réellement notre difficulté. Ou en tout cas, il recèle une ambiguïté fondamentale. Car il peut vouloir dire deux choses. D'une part il peut vouloir dire que les explications programmatiques, qui sont pertinentes causalement mais qui n'indiquent pas des causes efficaces, sont toujours en attente, en quelque sorte, d'explications par des processus effectifs. Il laisse donc en ce sens entendre que la causalité réelle intervient au niveau microphysique, qu'il y a une causalité primitive au niveau physique, qui agit comme une force réelle active et efficace. Ou bien le modèle en question ne laisse rien entendre de tel et dit en fait seulement que les propriétés pertinentes causalement au niveau supérieur « pointent vers », ou indiquent, celles du niveau inférieur, sans qu'on entende par là qu'il soit illégitime de faire des attributions causales authentiques au niveau supérieur.

Si nous adoptons la première hypothèse, comment éviter, tout d'abord, que les propriétés « causalement pertinentes » selon notre modèle, mais non pas causalement efficaces, ne paraissent avoir un statut causal dérivé, ou étiole, par rapport à celles qui le sont vraiment ? Auquel cas, comme dans les conceptions envisagées précédemment, on en reviendra à la suspicion qu'elles sont causalement épiphénoménales. Mais ensuite et surtout, même si l'on accepte cette conséquence, quelle raison avons-nous de penser qu'il existe une causalité brute microphysique qui fait, en quelque sorte, tout le travail dans la salle des machines ? Pourquoi supposer qu'il y a un niveau physique primitif ? Prenons l'exemple paradigmatique de réduction, celui de la température à l'énergie cinétique moyenne des molécules. L'objection de la réalisabilité multiple est que, pour les propriétés mentales, à la différence de la température, ces propriétés peuvent être réalisées différemment dans un organisme et dans un autre, et dans la composition physique d'un même organisme. Mais la température elle-même est réalisée différemment dans un type de corps et dans un autre. La température d'un gaz est une chose, mais la température d'un plasma n'est pas la même chose, parce qu'un plasma n'est pas constitué de molécules mais de particules subatomiques ; la température des solides est encore quelque chose de différent, et la température de l'espace vide est encore autre chose. Devons-nous en conclure que le concept de température est épiphénoménal, comme dans nos exemples les propriétés dispositionnelles, les propriétés mentales et un ensemble d'autres propriétés de niveau « supérieur » ? Non, car la physique peut



expliquer comment les liquides, les solides et l'espace peuvent être à la même température. On dit que si deux corps sont en équilibre thermique l'un avec l'autre, et si l'un est en équilibre thermique avec un troisième, alors le second est aussi en équilibre avec le troisième. En d'autres termes on peut former une relation d'équivalence entre des systèmes en équilibre thermique. Mais nous n'avons pas pour autant besoin d'identifier la température avec ses réalisations pour autant. Cela s'accorde apparemment fort bien avec la notion invoquée dans le modèle précédent d'une propriété qui « programme » un certain nombre d'événements. La température est une propriété qui fait abstraction de ces différences en réalisations physiques. Mais en ce sens la température figurera dans des explications programmatiques, et elle sera, de ce point de vue, de niveau « supérieur ». Fort bien. Mais qu'est-ce qui nous autorise alors à dire qu'il existe un niveau physique de base, primitif, et où opère la causalité ? Qu'est-ce qui nous interdit de supposer que les explications physiques sont des explications programmatiques elles-mêmes, sans que l'on puisse jamais localiser une causalité « efficace », un processus causal *noté* ? Mais si on adopte la première interprétation du modèle proposé, il n'y a aucune raison de supposer le contraire. Après tout, il pourrait y avoir, indéfiniment, des dispositions au niveau physique, sans que jamais on atteigne leur « base catégorielle », le niveau plancher qui les conditionne.

On doit donc adopter la seconde interprétation et dire que les propriétés programmatiques au lieu de présupposer une causalité efficace de niveau inférieur sont simplement des désignations, ou des indications de propriétés inférieures, sans que l'on sache où l'on devra s'arrêter dans la descente dans les niveaux. Mais si nous admettons cela, que devient la thèse initiale du physicalisme non réductionniste, la thèse selon laquelle si Dieu crée un monde dans ses caractéristiques physiques, il a créé du même coup toutes les autres caractéristiques ? Cette supposition ne cesse pas d'être correcte, puisque le modèle programmatique fait référence à cette dernière des niveaux physiques. Mais elle n'a aucun pouvoir explicatif.

Elle n'a aucun pouvoir explicatif, parce que s'il est supposé, comme le veut le physicalisme non réductionniste, que les explications causales de niveau supérieur sont des explications causales *bona fide*, mais qu'elles ne font que spécifier des processus causaux inférieurs, nous ne comprenons toujours pas quelle est la relation entre les deux types de niveaux. Que l'on soutienne ou non que les propriétés causales explicatives au niveau supérieur figurent dans des explications causales en vertu de lois, le modèle de l'explication

programmatique présuppose bien qu'il existe des lois au niveau microphysique inférieur. C'est en vertu de ces lois que nous pouvons « ne pas avoir de doutes » sur la pertinence causale de propriétés N que nous voulons par ailleurs décrire à un autre niveau au moyen de propriétés P. Mais si nous admettons la thèse de la réalisabilité multiple, celle de l'identité des événements particuliers de type P et des événements particuliers de type N, comment pouvons-nous admettre que les explications causales que nous donnons au niveau P puissent produire des corrélations entre des événements qui, chacun en particulier, tombent aussi sous des lois microphysiques quelconques ? Si nous disons qu'il existe des lois du mental, par quelle coïncidence deux événements qui se trouvent mis en relation par une loi mentale peuvent-ils se trouver mis en relation par des lois microphysiques sous lesquelles ces événements tombent également ? Et si nous n'admettons pas qu'il y ait des lois du mental, mais seulement des généralisations approximatives, voire des attributions causales purement singulières, il demeure tout aussi mystérieux qu'il puisse y avoir des lois physiques strictes qui sous-tendent ces généralisations ou attributions causales, sans qu'on puisse dire qu'il y ait la moindre connexion entre les deux niveaux. Sauf à dire que c'est une coïncidence ou un hasard, la rencontre de deux séries causales indépendantes, le physicaliste voudra pouvoir établir des relations entre les niveaux. Et si c'est le cas, je crains que cela ne vaille dire que l'idéal réductionniste a encore de beaux jours devant lui<sup>1</sup>.

Pascal ENGEL

Université de Caen, CREA, Ecole Polytechnique  
et Institut Universitaire de France.

1. Des versions de ce texte ont été lues à une réunion du groupe « Le Mercure philosophique » à Paris en novembre 1992, au Moral Sciences Club à Cambridge (décembre 1992), et à l'Université de Bristol en février 1993. Je remercie les participants pour les remarques, et en particulier Joëlle Proust, Tom Baldwin, Hugh Mellor, Andrew Woodfield et Adam Morton, ainsi que Philip Pettit.