

## Héritabilité causale et propriétés émergentes

Brian Garrett

Volume 27, Number 1, Spring 2000

Le matérialisme contemporain

URI: <https://id.erudit.org/iderudit/004924ar>

DOI: <https://doi.org/10.7202/004924ar>

[See table of contents](#)

Publisher(s)

Société de philosophie du Québec

ISSN

0316-2923 (print)

1492-1391 (digital)

[Explore this journal](#)

Cite this article

Garrett, B. (2000). Héritabilité causale et propriétés émergentes. *Philosophiques*, 27(1), 139–159. <https://doi.org/10.7202/004924ar>

Article abstract

Jaegwon Kim has recently argued, using what he calls “The Causal Inheritance Principle”, that multiply realizable properties are not causal, scientific kinds. My primary concern is to reply to Kim's arguments against nonreductive physicalism. I shall defend the idea that there are, indeed, more causal powers than the physical. But this should come as no surprise, since there are more individuals than the total number of fundamental physical particles. Reflections on the nature of individuals, specifically, on their ability to survive through change or replacement of their parts, indicates that individuals have causal powers nonidentical with the causal powers of the physical individuals that constitute them. I claim that, if this is plausible, the rejection of The Causal Inheritance Principle is also plausible.

# Héritabilité causale et propriétés émergentes

BRIAN GARRETT

bgarrett@yorku.ca

Department of Philosophy

York University, Canada

**RÉSUMÉ.** — Sur la base de ce qu'il a appelé « le principe d'héritabilité causale », Jaegwon Kim a soutenu que les propriétés réalisables de façons multiples ne constituent pas des sortes causales scientifiques. Mon principal objectif est de répondre aux arguments de Kim contre le physicalisme non réductionniste. Je défends l'idée qu'il existe plus de pouvoirs causaux que les seuls pouvoirs causaux physiques. Cela n'a rien de surprenant puisqu'il existe plus de particuliers que le nombre total de particules physiques fondamentales. Et la réflexion sur la nature des individus, et plus spécifiquement sur leur capacité à préserver leur identité à travers le changement ou le remplacement de leurs parties, indique que les individus ont des pouvoirs causaux distincts de ceux des particuliers physiques qui les constituent. Je soutiens que si cela est plausible, alors l'abandon du principe d'héritabilité causale l'est tout autant.

**ABSTRACT.** — Jaegwon Kim has recently argued, using what he calls “The Causal Inheritance Principle”, that multiply realizable properties are not causal, scientific kinds. My primary concern is to reply to Kim’s arguments against nonreductive physicalism. I shall defend the idea that there are, indeed, more causal powers than the physical. But this should come as no surprise, since there are more individuals than the total number of fundamental physical particles. Reflections on the nature of individuals, specifically, on their ability to survive through change or replacement of their parts, indicates that individuals have causal powers nonidentical with the causal powers of the physical individuals that constitute them. I claim that, if this is plausible, the rejection of The Causal Inheritance Principle is also plausible.

## 1. Introduction : épiphénoménisme ou pouvoirs causaux émergents?

Le physicalisme non réductionniste pose que les particuliers, qu'il s'agisse d'objets ou d'événements, sont des entités physiques, tout en niant que chaque aspect ou propriété du monde soit identique à un aspect ou à une propriété physique du monde. Cette thèse est appuyée, notamment, par le fonctionnalisme qui veut que les propriétés mentales soient des propriétés fonctionnelles distinctes des propriétés physiques qui les réalisent de façon contingente. Récemment, le physicalisme non réductionniste a été sérieusement critiqué, et l'une des principales difficultés qui lui a été adressée est celle de savoir si les propriétés survenantes irréductibles<sup>1</sup> ont un rôle causal dans la

---

1. J'entends la notion de survenance (*supervenience*) en un sens qui implique une relation de dépendance et une relation de réalisabilité multiple : (D) Nécessairement, pour tout  $x$  et toute

nature ou si elles ne sont que de simples épiphénomènes, sortes d'appendices aux processus physiques que les matérialistes tiennent pour causalement responsables du changement. Pour plusieurs, la conséquence épiphénoméniste semble inacceptable dans la mesure où elle minerait certaines conceptions plausibles de l'action intentionnelle, de la connaissance et de la responsabilité. Mais si ces propriétés irréductibles (émergentes?) ont réellement une efficacité causale, alors il semble que cela reviendrait à accepter l'existence de pouvoirs causaux qui échappent à la théorie physique, ce qui violerait le principe de clôture du domaine physique, à savoir que chaque changement physique a une explication physique complète, ou une cause entièrement physique. Il semble donc que le physicalisme non réductionniste soit confronté à un dilemme : ou bien les propriétés irréductibles sont causalement inertes et ne contribuent en rien à l'explication des changements physiques, ou bien elles sont causalement efficaces et la théorie physique est incomplète.

Dans cet article, je discute deux arguments de Jaegwon Kim, le plus tenace des critiques du physicalisme non réductionniste. Selon Kim, les deux termes du dilemme devraient être rejetés. Il propose de revenir plutôt à une forme de physicalisme réductionniste, à savoir la thèse de l'identité type-type. Dans ce qui suit, je soutiens que le point de vue de Kim nous engage à une ontologie très peu plausible, et qu'il a certaines implications gênantes relativement à la question de l'identité des personnes. Mais mon principal objectif est de répondre à certains arguments importants que Kim a dirigés contre la thèse du physicalisme non réductionniste. Cependant, plutôt que de défendre cette thèse telle qu'elle est généralement entendue, c'est-à-dire comme impliquant la thèse de l'identité token-token, mes arguments en suggèrent une version qui rejette cette autre thèse, à laquelle le physicalisme non réductionniste est généralement associé. J'endors le second terme du dilemme et défends l'idée qu'il existe plus de pouvoirs causaux que ce que nous fournit le monde physique<sup>2</sup>. Or cela n'a rien de surprenant, dans la mesure où il est tout à fait plausible qu'il existe plus de particuliers que le nombre total de particules physiques fondamentales plus les sommes météorologiques que constituent les agrégats physiques. Si un particulier, à un temps donné, n'est pas identique à la somme météorologique de ses parties physiques constituantes<sup>3</sup>, alors la différence entre le particulier et celle-ci doit tenir à certaines propriétés. Autrement dit, si deux particuliers qui ont toutes leurs

---

propriété  $M$  de  $x$ ,  $x$  a une propriété physique  $P$  telle que nécessairement tout  $x$  qui a la propriété  $P$  a la propriété  $M$ ; ( $RM$ ) Nécessairement, pour toute propriété mentale  $M$  et pour toute propriété physique  $P$ , si  $P$  nécessite  $M$ , il est possible qu'un objet possède  $M$  mais ne possède pas  $P$ .

2. Kim distingue deux thèses : l'émergentisme et la thèse de la réalisabilité physique (*physical realizationism*). Et sur la base du principe de l'héritabilité causale, il soutient que cette dernière serait une thèse moniste. Mon principal argument est que la réalisabilité physique n'est pas moniste, car le principe de l'héritabilité causale est erroné.

3. Cela suppose le rejet du principe d'extensionnalité météorologique qu'on peut exprimer de la façon suivante : Pour tout  $x$ , pour tout  $y$  et pour toute partie  $p$ , si certaines parties composent

parties physiques en commun sont différents, ce doit être en vertu d'une différence dans certaines de leurs propriétés. Et, à mon avis, ces propriétés sont des propriétés causales.

La réflexion sur la nature des individus, et plus spécifiquement sur leur capacité à préserver leur identité à travers le changement ou le remplacement de certaines de leurs parties, indique que les pouvoirs causaux de certains individus (à un temps  $t$ ), ne sont pas identiques aux pouvoirs causaux de leurs parties physiques constituantes considérées dans leur totalité (au temps  $t$ ). Considérez par exemple la célèbre statue de Myron, le Discobole, et la masse de bronze qui le compose. Si le Discobole peut survivre au remplacement de certaines de ses parties, alors que la masse de bronze ne le peut pas, ou si la masse de bronze peut changer de forme de façon pratiquement indéfinie tout en préservant son identité, alors que le Discobole ne le peut pas, alors ce serait une raison de nier que le Discobole est identique à la masse de bronze qui le compose (au temps  $t$ ). Cette dernière compose le Discobole mais la composition n'est pas l'identité — du moins selon ce point de vue qui nie le principe d'extensionnalité méréologique. Cet argument métaphysique est connu et, bien qu'il porte à controverse, c'est un argument que j'accepte. Je veux soutenir que ces individus sont numériquement distincts parce qu'ils possèdent des propriétés différentes, et que certaines de ces propriétés sont justement des propriétés causalement pertinentes ou efficaces. Si le Discobole et la masse de bronze qui le compose ont (au temps  $t$ ) exactement les mêmes propriétés physiques, alors leur différence doit s'expliquer par une différence de propriétés qui ne sont pas physiques mais qui ne sont pas pour autant épiphénoménales. Voilà donc une problématique où se rencontrent la métaphysique de l'esprit et celle du changement, de la composition et de l'identité.

Mais, qu'en est-il du principe de clôture de la théorie physique qui veut que tout changement physique ait une explication ou une cause physique? Selon mon point de vue, l'existence d'objets qui sont *composés* de sommes d'objets physiques sans leur être identiques ne viole pas, à strictement parler, le principe de clôture du domaine physique. Cela permet simplement de disposer de ressources ontologiques *additionnelles* à celles dont dispose la théorie physique, ressources qui peuvent servir dans des explications causales additionnelles. Si nous acceptons dans notre ontologie des propriétés causales distinctes des propriétés physiques, nous ne devons pas pour autant nier que chaque événement physique, ou chaque exemplification physique de propriétés, ait une cause physique ou une explication physique. Mais cela suggère qu'il existe, en plus des explications physiques, d'autres explications

---

l'objet  $x$  au temps  $t$  et ces mêmes parties composent l'objet  $y$  au temps  $t$ , alors  $x = y$ . Pour des arguments récents contre ce principe, voir Thomson, J., « The Statue and the Clay », *Nous*, 1998, p. 149-172 ; Baker, L. R., « Why Constitution Is Not Identity », *Journal of Philosophy*, 94, 1997, p. 599-621 ; Doepke, F., « Spatially Coinciding Objects », *Ratio*, 24, 1982.

causales de certains événements physiques : des explications qui réfèrent à des propriétés et à des événements causaux non physiques. À mon avis, l'existence de propriétés causales non physiques est compatible avec le principe de clôture de la théorie physique dans le sens suivant. D'abord, je ne nie pas que chaque événement physique ou chaque exemplification de propriétés ait une cause physique, mais l'existence de propriétés causales qui ne sont pas des propriétés physiques est incompatible avec une doctrine différente, bien que similaire, que j'appelle « la clôture causale de la théorie physique » et selon laquelle il n'existerait pas de propriétés ou d'événements causaux autres que les propriétés ou les événements physiques. C'est cette doctrine que je considère être fausse<sup>4</sup>. Les événements mentaux ne sont pas numériquement identiques aux événements physiques qui les composent, mais ils ont des relations causales avec d'autres événements physiques, tout comme les événements physiques qui composent un événement mental ont des relations causales avec d'autres événements<sup>5</sup>.

Évidemment, nous nous trouvons sur un terrain métaphysique glissant, car nous jonglons simultanément avec les notions de propriétés, d'objets, et d'événements. Ma discussion présuppose que les événements sont des particuliers et que, en tant que tel, il n'est pas incohérent de se poser des questions quant à leur composition par d'autres événements. Le point de vue que je tente de défendre dans ce qui suit est donc le suivant : les événements physiques composent les événements mentaux mais, de façon analogue au Discobole et à la masse de bronze qui le compose, la relation de composition n'est pas la relation d'identité. L'événement mental et les événements physiques qui le composent ont en commun exactement les mêmes propriétés physiques, mais l'événement mental n'en est pas moins distinct de la somme des événements physiques qui le composent. De plus, les propriétés qui distinguent l'événement mental de la somme des propriétés physiques qui le composent doivent être entendues comme étant causalement pertinentes.

Pour défendre ce point de vue, je dois montrer (i) que les arguments à l'appui d'une telle différence d'identité numérique d'objets constitués des mêmes parties physiques sont plausibles, (ii) que les propriétés en vertu desquelles un objet est différent de ses parties physiques constituantes ne sont pas des propriétés physiques, et (iii) que ces propriétés sont causalement pertinentes. Mais avant de développer ce point de vue, il est utile de montrer clairement en quoi il est différent du physicalisme de Kim. Aussi, dans la section suivante,

---

4. Le physicalisme non réductionniste que j'endors rejette l'identité token-token, mais il admet que les événements mentaux ont des propriétés physiques (ils ont des constituants physiques). Bien qu'ils ne soient pas numériquement identiques à des événements physiques, ils n'en ont pas moins des propriétés physiques. C'est en ce sens que mon point de vue est compatible avec la clôture du domaine physique.

5. Ce point de vue entraîne donc une forme de surdétermination. Mais celle-ci me semble acceptable si les causes numériquement distinctes sont dans la relation métaphysique de la partie au tout. Voir Garrett, B., « Non-Epiphenomenal Event Dualism », manuscrit.

je discute l'attaque de Kim contre le physicalisme non réductionniste et contre le fonctionnalisme. Je propose ensuite, dans une autre section, une critique du point de vue de Kim ainsi qu'un diagnostic des raisons qui l'ont mené à ses conclusions réductionnistes. Finalement, dans la dernière section, je tente de montrer en quoi les questions concernant la composition et l'identité des objets à travers le temps permettent d'éclairer l'argument de Kim. Si nous souscrivons à la conception métaphysique que je viens de décrire brièvement, nous pouvons rejeter le principe même sur lequel repose l'argument de Kim.

## 2. L'argument de Kim contre le physicalisme non réductionniste

Dans un texte récent<sup>6</sup>, Kim critique l'interprétation métaphysique du fonctionnalisme qui est largement admise par les tenants du physicalisme non réductionniste. Les propriétés fonctionnelles sont généralement conçues comme étant des propriétés de deuxième ordre. Dire qu'un système *S* a une certaine propriété fonctionnelle *F* revient à dire que *S* a certaines propriétés de premier ordre *P1* ou *P2* ou *P3*, etc., telles que le fait d'avoir l'une ou l'autre de ces dernières est causalement suffisant pour produire certains effets. Comme Kim le note :

*F* est une propriété de deuxième ordre définie sur un ensemble *B* de propriétés de base si et seulement si *F* est la propriété d'avoir une certaine propriété *P* appartenant à l'ensemble *B*, telle que *D(P)*, où *D* spécifie une certaine condition sur les membres de *B'*.

Si *F* est une propriété fonctionnelle, alors la condition *D* spécifie un certain rôle fonctionnel. Ainsi, par exemple, la solubilité est une propriété fonctionnelle de deuxième ordre. Pour qu'une substance, disons le sel, soit soluble, il doit exister une certaine propriété de premier ordre, disons la propriété d'être composé chimiquement de NaCl, telle que le fait d'avoir cette constitution chimique suffit à ce que le sel se dissolve s'il est placé dans un certain liquide non saturé approprié.

Il a été largement admis que le rôle fonctionnel est une propriété distincte des propriétés qui le réalisent. Par exemple, la solubilité est une propriété distincte de la propriété qui caractérise la constitution chimique du sel ou de celle qui caractérise le sucre : chacune de ces dernières réalise la solubilité. Cette idée a permis à certains fonctionnalistes comme Hilary Putnam, du moins à l'époque où il était encore fonctionnaliste, et Jerry Fodor de soutenir un certain dualisme des propriétés, puisqu'ils croyaient que les propriétés fonctionnelles ne sont pas des propriétés physiques, compte tenu de la thèse de la réalisabilité multiple qui veut qu'une même propriété fonctionnelle puisse avoir des réali-

6. Kim, J., « The Mind-Body Problem : Taking Stock After Forty Years », dans Tomberlin, J., dir., *Philosophical Perspectives*, vol. 11, Boston, Blackwell, 1997.

7. *Ibid.*, p. 195.

sations physiquement hétérogènes, comme le sel et le sucre dans l'exemple de la solubilité<sup>8</sup>. De plus, comme les propriétés fonctionnelles sont essentiellement caractérisées en faisant référence à des propriétés de premier ordre qui ont un certain rôle causal, le fonctionnalisme est apparu comme une façon de résoudre facilement les problèmes traditionnels concernant l'efficacité causale des propriétés mentales, désormais considérées comme des propriétés fonctionnelles. En effet, si les propriétés mentales sont des propriétés fonctionnelles de deuxième ordre, alors elles seraient causalement pertinentes, par définition, puisque les propriétés fonctionnelles requièrent l'existence de propriétés de premier ordre qui satisfont un certain rôle causal<sup>9</sup>. Finalement, un des attraits du fonctionnalisme en philosophie de l'esprit est de s'inscrire tout naturellement dans une certaine ontologie fonctionnaliste. Autrement dit, cette approche fournirait le schéma d'une ontologie générale : les propriétés fonctionnelles, comme William Lycan l'a suggéré, se retrouveraient à divers niveaux dans une organisation hiérarchique de la nature :

Si nous considérons la nature comme étant hiérarchiquement organisée de cette façon, alors la distinction « fonction/structure » devient relative : quelque chose est un rôle par opposition à l'occupant d'un rôle, un état fonctionnel par opposition à une réalisation de cet état, ou vice versa, modulo un certain niveau spécifique de la nature<sup>10</sup>.

Ainsi, la nature comprendrait une hiérarchie de propriétés ontologiquement distinctes et causalement efficaces, et les propriétés se situant à un certain niveau seraient survenantes à celles se situant à un niveau inférieur. Ultimement, toutes seraient survenantes à des propriétés fondamentales, supposément les propriétés physiques de l'univers. La question de savoir si une propriété est ou non une propriété fonctionnelle de deuxième ordre serait une question pragmatique qui dépend de nos intérêts dans divers contextes<sup>11</sup>. Relativement à la psychologie, les propriétés neurologiques

---

8. Voir Putnam, H., « Philosophy and Our Mental Life », dans *Mind, Language and Reality : Philosophical Papers*, vol. 2, Cambridge (Mass.), Cambridge University Press, 1975 ; et Fodor, J., « Special Sciences : Still Autonomous After All These Years », dans Tomberlin, J., dir., *Philosophical Perspectives*.

9. Cependant, Kim lui-même ainsi que Ned Block ont soutenu que la propriété d'avoir une propriété causalement efficace n'est pas elle-même une propriété causalement efficace ; voir Block, N., « Can the Mind Change the World », dans MacDonald, C. et Macdonald, G., dir., *Philosophy of Psychology. Debates on Psychological Explanation*, Oxford, Blackwell, 1995.

10. Lycan, W., « The Continuity of Levels of Nature », dans Lycan, W., dir., *Mind and Cognition*, Cambridge (Mass.), Blackwell, 1990, p. 78.

11. Cette façon de parler pourrait laisser croire que les propriétés fonctionnelles et les propriétés de deuxième ordre sont une seule et même chose, mais ce n'est pas le cas, comme l'a noté Fodor dans « Special Sciences : Still Autonomous After All These Years ». Le jade est une propriété de deuxième ordre car ses exemplifications requièrent l'exemplification de propriétés réalisantes, mais elle n'est pas fonctionnelle car elle requiert des propriétés réalisantes très spécifiques et non pas simplement toute propriété réalisante qui capturerait les causes et les effets macrophysiques du jade.

seraient des propriétés réalisantes de premier ordre, mais relativement à la sociologie, les propriétés psychologiques seraient elles-mêmes des propriétés réalisantes de premier ordre.

Jaegwon Kim a récemment critiqué l'approche fonctionnaliste en s'appuyant sur divers arguments. Le premier, et peut-être le plus important, est dirigé contre les arguments de réalisabilité multiple (RM) qui appuient l'indépendance ontologique des propriétés fonctionnelles. Kim tente de bloquer ces arguments dans le but de montrer que si les propriétés survenantes peuvent être réalisées de multiples façons, alors elles ne peuvent être des propriétés scientifiques ou naturelles légitimes, ce qui serait une mauvaise nouvelle pour les tenants de l'irréductibilité ontologique des sciences « spéciales ». À tout le moins, ce serait une mauvaise nouvelle pour ceux qui, comme Jerry Fodor, soutiennent que les propriétés irréductibles pourraient être étudiées systématiquement dans le contexte de lois scientifiques exprimant leurs pouvoirs causaux<sup>12</sup>.

Le deuxième argument de Kim s'attaque à la définition des propriétés fonctionnelles. Si celles-ci doivent être conçues comme des propriétés de deuxième ordre, alors nous n'aurions pas de motifs pour faire des affirmations ontologiques substantielles à leur propos. Si une propriété fonctionnelle est une propriété de deuxième ordre, c'est-à-dire la propriété d'avoir certaines propriétés de premier ordre satisfaisant certains rôles causaux, alors, nous dit Kim, une telle propriété ne serait pas réellement distincte des propriétés de premier ordre qui la réalisent. Nous discutons cet argument de façon plus détaillée dans ce qui suit.

La thèse de la réalisabilité multiple a été utilisée par Putnam et par Fodor pour soutenir que les propriétés physiques et mentales sont ontologiquement distinctes. L'idée est tout simplement qu'une propriété mentale serait réalisée par une longue disjonction, voire une disjonction infinie, de sortes physiques. Or une longue disjonction de sortes physiques n'est pas elle-même une sorte physique, car le prédicat disjonctif qui dénote cette « propriété » n'aurait pas le caractère de projectivité qui est un des critères généralement admis pour qu'un prédicat dénote une sorte physique. Ainsi, la propriété mentale ne pourrait être identifiée à une propriété physique, puisque cette dernière n'existe tout simplement pas. Kim renverse cet argument et il l'utilise pour montrer justement que si les propriétés mentales peuvent être réalisées de multiples façons, alors elles ne constituent tout simplement pas des sortes naturelles, dans la mesure où les « propriétés » disjonctives qui leurs sont coextensives ne sont pas des sortes naturelles. Le point central de l'argument de Kim est que les propriétés survenantes doivent hériter leurs

---

12. Fodor, J., « Special Sciences, or the Disunity of Science as a Working Hypothesis », dans Block, N., dir., *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980, p. 120-133.



pouvoirs causaux, et par conséquent leur statut taxinomique, des propriétés physiques qui les réalisent. Considérons les trois principes suivants :

*Principe de réalisabilité multiple* (RM) : Nécessairement, pour toute propriété mentale  $M$  et pour toute propriété physique  $P$  qui nécessite  $M$ , il est possible qu'un objet possède  $M$  mais ne possède pas  $P$ .

*Principe d'héritabilité causale* (HC) : Si la propriété  $M$  est réalisée dans un système, au temps  $t$ , par une propriété physique  $P$ , les pouvoirs causaux de cette occurrence particulière de  $M$  (disons  $Mt$ ) sont identiques aux pouvoirs causaux de cette occurrence particulière de  $P$  (disons  $Pt$ ).

*Principe d'individuation causale des sortes naturelles* (ICS) : Les sortes naturelles en science sont individuées sur la base de leurs pouvoirs causaux ; c'est-à-dire que les objets et les événements sont d'une même sorte naturelle ou partagent une même propriété seulement s'ils ont des pouvoirs causaux similaires<sup>13</sup>.

Il est utile de commenter brièvement chacun de ces principes. Selon (RM), les propriétés mentales sont réalisées, et peuvent être réalisées, par des propriétés physiques hétérogènes. Les exemples standards incluent la possibilité que le cerveau d'un extraterrestre, un « cerveau » informatique de silicone et un cerveau humain puissent réaliser la même propriété psychologique, par exemple la propriété de croire-que-Leibniz-était-brillant. Le troisième principe illustre le point de vue de Kim à propos des sortes naturelles. En fait, il exprime sans doute deux thèses différentes, à savoir : (ICS.i) les sortes naturelles en science sont des sortes causales, et (ICS.ii) si  $K = K^*$ , alors  $K(a)$  doit être causalement similaire à  $K^*(b)$ <sup>14</sup>. (ICS.i) paraît douteux s'il est entendu comme une généralisation universelle, puisque Kim ne veut probablement pas dire que les sortes en mathématiques ou dans les sciences abstraites sont des sortes causales<sup>15</sup>. Cela laisse ouverte la possibilité que la psychologie elle-même soit une science abstraite, c'est-à-dire une science qui s'intéresserait essentiellement à la rationalité, et dont l'objectif serait de produire des rationalisations plutôt que des généralisations causales. Peut-être. Mais une telle objection à (ICS.i) ne peut être utilisée dans le cas d'autres sciences « spéciales » qui s'intéressent également à des propriétés survenantes. Il est en effet difficile de croire que les propriétés géologiques ou biologiques sont des propriétés abstraites causalement inefficaces. Donc j'accepte (ICS.i).

(ICS.ii) porte sur l'individuation des sortes scientifiques. Cependant, ce principe est plutôt vague : jusqu'à quel point deux occurrences d'une pro-

13. Fodor, J., « Methodological Solipsism », dans *RePresentations. Philosophical Essay on the Foundations of Cognitive Science*, Cambridge (Mass.), MIT Press, 1981, p. 225-253.

14. Où «  $a$  » et «  $b$  » réfèrent à des particuliers.

15. Mais pourquoi pas? Si les propriétés sont des sortes causales, dans la mesure où leurs exemplifications en sont des réalisations physiques, alors puisque les propriétés mathématiques sont exemplifiées par des états physiques, celles-ci seraient également des sortes causales. Mais peut-être qu'il ne suffit pas qu'une propriété ait des réalisations physiques pour qu'elle constitue une sorte causale ; quelque chose de plus est peut-être requis.

priété (dans des objets distincts) doivent-elles être causalement similaires pour qu'elles soient des occurrences distinctes d'une même propriété? Les jugements de similarité et d'hétérogénéité sont évidemment relatifs à un arrière-plan contextuel, et cela vaut également pour les jugements de similarité causale. Le problème n'est pas tant que (ICS.ii) soit erroné (je l'endosse également), mais qu'il puisse être interprété de telle sorte qu'il constituerait une pétition de principe contre le physicalisme non réductionniste. Si on exige que les jugements de similarité causale utilisent le vocabulaire physicaliste, alors le principe devient douteux : les tenants de la thèse selon laquelle il existe des propriétés causalement efficaces mais irréductibles à des propriétés physiques n'accepteraient pas cette interprétation. Si aucune contrainte de ce type ne restreint la notion de similarité causale, alors le principe semble difficilement contestable. Cependant, comme je le montre dans la section suivante, l'argument de Kim devient alors douteux.

Finalement, selon le principe d'héritabilité causale, les occurrences de propriétés réalisées de multiples façons hériteraient leurs propriétés causales directement des occurrences de propriétés physiques qui les réalisent. Ce principe est bien évidemment une conséquence directe de la thèse de l'identité token-token. Si une occurrence de *Mi*, la croyance-que-Leibniz-était-brillant, est identique à un certain événement neuronal, disons *Ni*, alors cette occurrence mentale doit avoir exactement les propriétés causales de l'événement neuronal auquel elle est identique! Aussi, dans la mesure où on est enclin à accepter la thèse de l'identité token-token, on devrait accepter (HC), puisque selon ce point de vue moniste les occurrences de propriétés mentales sont identiques à des occurrences de propriétés physiques (même si les propriétés mentales elles-mêmes, ou types mentaux, ne sont pas identiques à des propriétés physiques). Ce point de vue s'accommode bien du monisme anomal de Davidson, pour qui les événements mentaux sont identiques à des événements physiques qui sont eux-mêmes les causes du comportement<sup>16</sup>. Bref, le principe d'héritabilité causale semble être endossé par les tenants du physicalisme non réductionniste, et par conséquent il devrait être sacro-saint.

Or ces trois principes semblent avoir des conséquences dévastatrices pour l'ontologie du physicalisme non réductionniste. Considérons la propriété mentale *M*, disons croire-que-*p*, et deux occurrences de cette propriété, *Mi* et *Mii*, réalisées chez deux individus, disons un extraterrestre et un humain. Selon (RM), *M* est réalisée par des événements physiques distincts, qui exemplifient également deux propriétés physiques *P1* et *P2*. Selon l'héritabilité causale, *Mi* et *Mii* auraient respectivement les mêmes propriétés causales que les deux événements physiques (*P1i* et *P2i*) qui les réalisent. Or, il semble y avoir un sérieux problème lorsque nous considérons le troisième principe. En effet, si nous supposons que *Mi* et *Mii* ont, respectivement, les mêmes propriétés causales que

---

16. Voir Davidson, D., « Mental Events », dans Foster, L. et Swanson, J. W., dir., *Experience and Theory*, Amherst, The University of Massachusetts Press, 1970.

*P1i* et *P2i*, la question se pose de savoir comment *Mi* et *Mii* devraient être individuées, c'est-à-dire sous quelle(s) sorte(s) naturelle(s) nous devrions les catégoriser. *Mi* et *Mii* auraient des pouvoirs causaux différents puisqu'elles auraient exactement les pouvoirs causaux des occurrences de propriétés physiques distinctes (*P1* et *P2*). Ainsi, compte tenu du principe d'individuation, *Mi* et *Mii* ne seraient pas deux occurrences d'une même sorte scientifique *M*, car les deux réalisations de *M* sont *causalement* diverses et, selon (ICS), elles ne pourraient être catégorisées comme appartenant à la même sorte naturelle. Donc, *M* ne peut constituer une sorte scientifique. Si (RM) vaut généralement pour les propriétés mentales, géologiques, météorologiques, etc., alors celles-ci ne constitueraient tout simplement pas des sortes scientifiques! Les prédicats utilisés dans ces disciplines exprimeraient simplement des concepts plutôt que des propriétés naturelles. Comme le note Kim :

Le raisonnement est simple : les occurrences de *M* qui sont réalisées par une même base physique doivent être groupées sous une même sorte puisque, comme nous le supposons, la base physique constitue une sorte causale ; et les occurrences de *M* qui ont des bases physiques différentes doivent être groupées sous des sortes différentes puisque, comme nous le supposons toujours, ces bases de réalisation physiques distinctes constituent des sortes causales distinctes. Étant donné que les sortes mentales sont réalisées par des sortes physiques causalement distinctes, il en résulte donc que les sortes mentales ne sont pas des sortes causales, et ainsi elles ne constituent pas des sortes scientifiques adéquates<sup>17</sup>.

L'exemple utilisé par Kim illustre bien son argument. Comme nous le savons, le jade est coextensif avec la jadéite et la néphrite, qui sont des sortes naturelles distinctes, ce que révèle leur description moléculaire. Une loi à propos du jade, comme par exemple « Le jade est vert », est nomologiquement coextensive à deux lois : « la jadéite est verte » et « la néphrite est verte ». Ainsi, « Le jade est vert » n'est pas une généralisation nomologique car le terme « jade » n'est pas projectible, puisque le jade est tantôt réalisé par la jadéite, tantôt par la néphrite, lesquelles sont chimiquement hétérogènes. Si nous essayons de confirmer l'énoncé général « Le jade est vert » et que tous nos échantillons sont des échantillons de jadéite, alors il ne serait pas légitime d'inférer sur cette base inductive que le prochain échantillon de jade, disons un échantillon de néphrite, sera probablement vert. Un tel saut inductif ne serait pas fondé compte tenu de la différence moléculaire entre la jadéite et la néphrite.

Si (RM) est vrai, alors les propriétés réalisées de multiples façons ne sont pas des sortes naturelles, tout comme le jade n'est pas une sorte naturelle compte tenu de ses réalisations hétérogènes. Cependant, le prédicat « jade » n'en est pas moins utile pour nos explications de la vie de tous les jours. Ainsi, selon Kim, ce prédicat exprimerait un concept plutôt qu'une propriété ayant une efficacité causale. Pour reprendre l'exemple que nous

---

17. Kim, J., « Multiple Realization and the Metaphysics of Reduction », *Philosophy and Phenomenological Research*, 52, 1992, p. 19.

avons déjà utilisé, croire-que-Leibniz-était-brillant ne serait pas une sorte naturelle, car cette propriété est réalisable de multiples façons par des propriétés physiques différentes, par exemple par des machines de silicone, par des extraterrestres, par des humains et ainsi de suite.

Le deuxième argument de Kim procède de la façon suivante. Si les propriétés fonctionnelles sont des propriétés de deuxième ordre, alors elles ne sont rien de plus que les propriétés de premier ordre qui les réalisent. Une propriété de deuxième ordre est formée par quantification existentielle sur un domaine de propriétés de premier ordre. La solubilité consiste dans le fait d'avoir une propriété (chimique) de premier ordre qui a certains effets appropriés (si une substance soluble est plongée dans un liquide approprié dans certaines conditions  $C$ , alors elle se dissout). Le jade est une propriété de deuxième ordre dans le sens où  $x$  est du jade s'il existe certaines propriétés de base qui satisfont le rôle causal du jade, à savoir la propriété d'être de la jadéite ou celle d'être de la néphrite. Mais comme le note Kim, ce serait pure magie si une telle quantification existentielle sur des propriétés de premier ordre suffisait à générer une nouvelle propriété distincte :

Nous pouvons commencer par reconnaître explicitement qu'en généralisant existentiellement sur un certain ensemble de propriétés, nous n'amenons pas à l'existence un nouvel ensemble de propriétés. Cela serait pure magie : par des opérations logiques sur notre notation, comme la quantification, nous ne pouvons modifier notre ontologie — nous ne pouvons ni la diminuer ni l'augmenter. [...] En quantifiant sur des propriétés, nous ne pouvons créer de nouvelles propriétés, pas plus que nous ne pouvons créer de nouveaux individus en quantifiant sur des individus. Quelqu'un a assassiné Dupont et cet assassin est ou Dumont, ou Durand, ou Tremblay. Ce « quelqu'un » qui a assassiné Dupont n'est pas une autre personne en plus de Dumont, Durand, et Tremblay, et il serait absurde de postuler une personne disjunctive, soit Dumont-ou-Durand-ou-Tremblay qu'on identifierait à l'assassin. Le même raisonnement vaut pour les propriétés de second ordre et leurs réalisations<sup>18</sup>.

Tout comme une disjonction de propriétés  $P1$  ou  $P2$  ou ...  $Pn$  n'exprime pas une propriété, une quantification existentielle comme « Il y a une propriété  $P$  telle que  $x$  est  $P$  » n'exprimerait pas non plus une propriété. Le fonctionnalisme, qui veut que les propriétés fonctionnelles soient des propriétés de deuxième ordre formées par quantification existentielle sur des propriétés de premier ordre, ne nous permettrait pas de réifier les propriétés fonctionnelles comme étant de nouvelles propriétés irréductibles. Par conséquent, si les propriétés invoquées dans les sciences spéciales sont bien des propriétés de deuxième ordre, elles ne seraient finalement pas différentes des propriétés de premier ordre qui les réalisent. Si, par exemple, la propriété d'être un glacier n'est rien d'autre que la propriété d'avoir certaines propriétés de premier ordre ayant un certain rôle causal que l'on peut attendre des glaciers — comme celle

---

18. Kim, « The Mind-Body Problem : Taking Stock After Forty Years », p. 201.

d'être de la glace et celle d'être de la neige —, alors la propriété d'être un glacier ne serait pas une propriété au-delà de ses propriétés réalisantes.

### 3. Objections à l'argument de Kim

Kim croit qu'il serait moins trompeur de parler de concepts plutôt que de propriétés de deuxième ordre. Il serait préférable, selon lui, d'entendre les propriétés fonctionnelles comme étant des concepts ou des descriptions qui dénotent des propriétés de premier ordre. Si les « propriétés » fonctionnelles sont entendues comme n'étant rien d'autre que des concepts, cela nous permettrait d'en préserver l'utilité conceptuelle tout en faisant une économie ontologique : nous pouvons utiliser de tels concepts fonctionnels tout en restant silencieux à propos des propriétés de premier ordre. Nous pouvons parler de la solubilité d'une substance sans avoir à être plus précis sur les propriétés chimiques qui sont suffisantes pour la solubilité. De tels concepts véhiculent de l'information pertinente dans les contextes où nous les utilisons, information qui pourrait être occultée si nous utilisions plutôt les prédicats exprimant les propriétés de premier ordre. L'approche de Kim est attrayante, principalement parce qu'elle est ontologiquement parcimonieuse : avoir une propriété fonctionnelle n'est rien d'autre que d'avoir une propriété réalisante quelconque. Nous n'avons pas besoin d'empiler les unes par-dessus les autres des propriétés de plusieurs niveaux différents.

Mais la position générale qui résulte de l'approche de Kim est difficile à croire, car si elle est correcte, alors il serait exclu que les propriétés fonctionnelles soient des sortes scientifiques. Plus exactement, il n'existerait pas de propriétés réalisables de multiples façons mais seulement des concepts qui peuvent être appliqués de multiples façons, et la hiérarchie de propriétés fonctionnelles irréductibles s'effondrerait. Les propriétés physiques existent et elles sont fondamentales mais elles ne seraient pas fondamentales au sens où elles seraient constitutives des propriétés des diverses sciences, mais simplement dans le sens où elles détermineraient les conditions d'applicabilité des concepts utilisés dans ces diverses sciences. S'il doit exister dans les sciences spéciales des propriétés ontologiquement robustes, alors celles-ci ne pourraient être réalisées de multiples façons, car elles seraient coextensives à des sortes causales légitimes du domaine physique. Dans la mesure où la plupart des « propriétés » des sciences spéciales sont réalisées de façons multiples, comme Jerry Fodor l'a souligné, l'argument de Kim serait du coup un argument pour l'élimination ontologique de la très vaste majorité des propriétés des sciences spéciales. Dès lors, l'extraterrestre et l'humain qui ont la croyance-que-Leibniz-était-brillant n'auraient plus de propriété en commun, bien que le *concept* « croire que Leibniz est brillant » pourrait leur être appliqué, tout comme le concept « jade » peut être appliqué à deux quantités de jadéite et de néphrite. L'argument de Kim est peu crédible.

Il semble peu plausible qu'il n'existe pas de propriétés des sciences spéciales réalisées de multiples façons. S'il n'existe pas de telles sortes scientifiques, alors que dire des objets qui instancient ces « propriétés »? Si le prédicat « est une montagne » peut être réalisé de façons multiples alors, selon Kim, il n'y aurait pas de propriété causale d'être une montagne. Cela semble avoir pour conséquence qu'à proprement parler, il n'existerait pas un objet d'étude scientifique dont la caractéristique est d'avoir cette propriété. Mais une telle conclusion n'est-elle pas difficile à croire? Quel serait donc cet objet que les géologues ont essayé de comprendre? De plus, cette conclusion est parfaitement générale : il n'y aurait pas d'objets scientifiques individués par des propriétés réalisées de multiples façons (car il n'y a pas de telles propriétés), de sorte que la plupart des objets que nous acceptons dans le discours de tous les jours n'existeraient pas comme objets scientifiques. Mais ici nous devons être prudents : l'argument de Kim n'entraîne pas la conclusion qu'il n'existe pas de montagnes ou qu'il n'existe pas de créatures psychologiques. Par contre, ce qui résulte de son argument, c'est qu'il ne peut y avoir de science causale des propriétés psychologiques ou de science causale de la propriété d'être une montagne.

Si les propriétés psychologiques sont réalisées de multiples façons — disons dans des créatures de silicone et dans des créatures de carbone — alors ces « propriétés » ne seraient pas des sortes causales tombant dans le domaine d'étude de la science psychologique, puisque aucune science psychologique ne pourrait subsumer sous une même catégorie ces deux créatures. Une science psychologique causale pourrait subsumer sous une même catégorie les êtres humains, à condition que les réalisations physiques des propriétés psychologiques des humains ne soient pas hétérogènes. Même si les extraterrestres, les machines à base de silicone et les être humains ne tombent pas tous sous des sortes naturelles de la psychologie humaine, cela n'exclut pas la possibilité de diverses sciences psychologiques, par exemple la psychologie proprement humaine, la psychologie des extraterrestres, et la psychologie des machines. Cependant, il faut bien voir que même si ces diverses sciences portaient toutes le nom de « psychologie », ces sciences ne seraient pas liées entre elles comme différentes familles d'une même espèce, car elles ne s'intéresseraient aucunement à des propriétés psychologiques communes à des créatures de diverses sortes. Ce qui est gênant avec un tel point de vue c'est qu'il aurait le désavantage de nous faire perdre des généralisations qui paraissent subsumer correctement ces diverses créatures, comme Ned Block l'a souligné<sup>19</sup>.

Si nous devons accepter le point de vue de Kim, certaines thèses métaphysiques tout à fait plausibles concernant l'identité personnelle et la continuité devraient être abandonnées. Considérons, par exemple, un processus extrêmement lent par lequel les neurones et les connexions neuronales de votre cerveau seraient graduellement remplacés par des mécanismes artificiels capa-

---

19. Block, N., « Anti-Reductionism Slaps Back », dans Tomberlin, J., dir., *Philosophical Perspectives*, p. 107-132.

bles d'effectuer les mêmes transferts d'énergie et d'information que ceux effectués par vos neurones. Afin d'éviter la dégénérescence prévue de vos cellules nerveuses et de leurs fonctions, il n'est pas inconcevable que vos neurones puissent être remplacés de la sorte par des éléments artificiels ou par un matériau biologique artificiellement « cultivé », éléments qui seraient physiquement différents des neurones d'origine. Mon intuition est qu'un tel processus pourrait préserver intactes nos propriétés psychologiques. Comme le but pratique visé dans un tel scénario est justement de préserver le fonctionnement psychologique de l'individu, c'est uniquement la préservation des effets pertinents pour ce but qui est désirée. Les mécanismes artificiels, qu'ils soient biologiques ou non, pourraient eux-mêmes être constitués de propriétés physiques hétérogènes et ils auraient donc des pouvoirs causaux différents. Ces mécanismes pourraient avoir des effets tout à fait distincts s'ils étaient placés dans d'autres médiums, mais dans le contexte d'un cerveau ils pourraient bien accomplir les mêmes tâches. Le point important est qu'un tel processus de remplacement graduel des neurones pourrait réussir et, si c'était le cas, qu'il produirait un patient qui, du point de vue de la troisième personne, serait en parfaite santé psychologique.

Mais en raison du changement supposé dans la base de réalisation qui transforme une neurophysiologie humaine en l'une ou l'autre de ces « physiologies » artificielles, Kim doit nier que le patient exemplifie les mêmes propriétés psychologiques que celles qu'il exemplifiait avant l'opération. Selon son point de vue, une telle opération devrait échouer en principe si le but est de préserver la psychologie d'origine du patient. Kim pourrait sans doute répondre que même s'il n'y a pas de continuité psychologique avant et après l'opération, il y aurait un isomorphisme fonctionnel. Mais cela équivaldrait à la perte de la personne en cours d'opération et, curieusement, ce serait une perte de propriétés psychologiques que seul un médecin pourrait détecter! À mon avis, même si cette objection n'est peut-être pas dirimante, dans la mesure où les opinions philosophiques à propos de la nature de l'identité personnelle divergent considérablement, elle soulève tout de même une sérieuse difficulté. Kim pourrait sans doute accepter la conséquence selon laquelle des opérations de ce genre ne préserveraient pas l'identité personnelle, puisqu'elles ne préserveraient pas la continuité psychologique. Mais les conséquences épistémologiques d'un tel point de vue sont quelque peu déroutantes. En effet, il semble bien que ni le patient d'une telle opération ni des observateurs extérieurs ne seraient en mesure de savoir qu'il y a eu perte de la personne au cours de l'opération, à moins qu'ils ne connaissent les faits physiques pertinents, ce qui semble épistémiquement contre-intuitif. Par ailleurs, Kim pourrait nier que l'identité personnelle dépende de la continuité et de la connectivité psychologiques<sup>20</sup>. Peut-être que l'identité personnelle à travers le temps pourrait être préservée en faisant l'économie de

---

20. Les notions de continuité psychologique et de connectivité psychologique ont été caractérisées par Derek Parfit. Selon Parfit, il y a connectivité psychologique lorsqu'il existe

ces deux caractéristiques. Cependant, je dois admettre que je ne comprends tout simplement pas en quel sens une théorie qui nie la nécessité de la continuité psychologique ou de la connectivité (plutôt que de nier simplement que ce sont là des conditions suffisantes) puisse être une théorie de l'identité des personnes, plutôt qu'une théorie de leur corps.

La position générale de Kim a donc des conséquences qui semblent difficilement acceptables. Qu'a-t-il bien pu se passer? La critique la plus intéressante (et sans doute la plus discutée) consiste à rejeter un ou plusieurs des trois principes sur lesquels repose l'argument de Kim. Dans ce qui suit, je propose de rejeter le principe d'héritabilité causale. Mais revenons d'abord à l'ambiguïté du principe d'individuation causale des sortes naturelles. Selon ce principe « [...] les objets et les événements sont d'une même sorte naturelle, ou partagent une même propriété, seulement s'ils ont des pouvoirs causaux similaires ». Mais si nous considérons deux particuliers quelconques, ceux-ci peuvent toujours être considérés comme ayant des pouvoirs causaux similaires ou hétérogènes, si nous prenons la peine de choisir de façon judicieuse la mesure de similarité. Ainsi, deux exemplifications d'une certaine propriété physique seront causalement hétérogènes relativement à une certaine mesure. Nos jugements de similarité causale dépendent des détails auxquels nous portons attention, c'est-à-dire des effets et des causes qui nous importent dans un contexte particulier.

L'objection est que nous devons toujours décider d'une certaine mesure de similarité avant de pouvoir évaluer la situation. Un tenant de l'efficacité causale des propriétés réalisables de multiples façons ne manquera pas de noter que, par définition, (RM) requiert l'hétérogénéité à un certain niveau de description et l'homogénéité à un autre niveau. Ainsi, les occurrences de propriétés survenantes réalisables de multiples façons ont des effets très similaires lorsque ces occurrences sont catégorisées à l'aide du vocabulaire de la discipline appropriée. On pourrait tenter de résister à cette objection en insistant sur le fait que les propriétés réalisables de multiples façons doivent être des sortes causales, que deux occurrences d'une même sorte doivent être causalement similaires, et qu'il serait correct de dire qu'elles le sont seulement si nous sommes forcés de mesurer leur similarité causale par des standards *physiquement* adéquats. Mais un tel raisonnement semble justement présupposer ce qu'il tente d'établir. Comme je l'ai déjà noté, la notion de « similarité causale » que nous utilisons dans cette discussion doit être neutre (*topic neutral*).

---

certaines connexions psychologiques directes entre des états mentaux du sujet. De telles connexions directes sont des relations causales. La connectivité psychologique est une question de degré qui dépend du nombre de telles connexions directes. Sur cette base, Parfit caractérise la notion de connectivité *forte* qui, bien qu'elle ne puisse être définie strictement, signifie tout simplement que le degré de connectivité psychologique est « assez grand » pour que le sujet soit une personne. La continuité psychologique est ensuite définie comme une chaîne continue de relations de connectivité *forte*, qui se chevauchent. Voir Parfit, D., *Reasons and Persons*, Oxford, Clarendon Press, p. 206.



On pourrait également être tenté de répliquer en notant que l'occurrence particulière d'une propriété survenante est un événement identique à l'occurrence de la propriété physique qui la réalise et, par conséquent, que le même critère de similarité causale devrait s'appliquer à « chacune ». Comme, selon (RM), les diverses réalisations sont hétérogènes, alors, étant donné l'identité token-token, les propriétés survenantes seraient aussi hétérogènes. Mais cette réplique passe à côté de l'objection. Même si nous concédions cette version du physicalisme non réductionniste — bien que finalement je désire résister à cette interprétation —, nous devrions de toutes façons poser la question de savoir si les deux événements sont ou non causalement similaires. Imaginons, par exemple, deux événements qui sont des cas d'ingestion d'une substance dormitive, dans un cas de l'opium et dans l'autre un sédatif obtenu à la pharmacie. Que ces deux événements soient causalement similaires ou non dépend, encore une fois, des types d'effets et de causes que nous considérons être pertinents. Si toutes les différences étaient pertinentes, alors nous devrions abandonner l'idée qu'une propriété causalement efficace puisse être exemplifiée plus d'une fois. Si nous ne considérons que la similarité de ce qu'on pourrait appeler les micro-effets comme étant pertinente, alors les deux événements pourraient bien être causalement différents, dans la mesure où il s'agit d'ingestion de deux types de sédatifs différents ; mais si la similarité d'effets macroscopiques sur la psychologie humaine est notre critère, ces événements pourraient être classifiés comme étant causalement similaires.

#### 4. Contre le principe d'héritabilité causale

Le principe d'héritabilité causale joue un rôle fondamental dans l'argument de Kim. Le nom de ce principe est particulièrement évocateur. Lorsque nous héritons de quelque chose nous obtenons quelque chose qui nous faisait défaut ; mais lorsque nous héritons, nous n'héritons pas nécessairement de la totalité des biens héréditaires : il peut arriver que nous n'obtenions qu'une partie du magot, le reste pouvant aller à des œuvres de charité ou à d'obscurs cousins oubliés. De plus, être un héritier ne requiert pas que nous soyons au préalable indigents. Or Kim utilise la notion d'héritabilité comme un principe qui n'accepte pas de degré et qui s'applique de façon *tout ou rien*. Les occurrences de propriétés mentales nettoient tout sur leur passage : elles héritent très exactement de tous les pouvoirs causaux des occurrences de propriétés physiques qui les réalisent. J'aimerais suggérer que la métaphore d'héritabilité ne devrait pas être restreinte de la sorte et que le principe devrait être révisé. Comme je l'ai déjà noté, le principe est le suivant :

*Principe d'héritabilité causale (HC) :* Si la propriété mentale *M* est réalisée dans un système, au temps *t*, en vertu d'une base de réalisation physique *P*, les pouvoirs causaux de cette occurrence particulière de *M* (disons *Mi*) sont identiques aux pouvoirs causaux de cette occurrence de *P* (disons *Pi*).

Cette thèse est particulièrement plausible pour une certaine version du physicalisme non réductionniste, à savoir la thèse du monisme anomal de Davidson. Si les événements mentaux sont des particuliers concrets et qu'ils sont identiques à des événements physiques, alors, en vertu de la relation d'identité<sup>21</sup>, ils doivent avoir exactement les mêmes propriétés que ces derniers, incluant les propriétés causales<sup>22</sup>. Mais le monisme anomal n'est pas la seule version possible du physicalisme non réductionniste. Celui-ci peut s'appuyer sur la notion de constitution, ou de « composition », qui est différente de celle d'identité<sup>23</sup>. Or, ce qui est vrai des particuliers, comme les objets, peut bien l'être également des événements<sup>24</sup>. Les événements peuvent être composés d'autres événements, mais si la composition n'est pas l'identité, alors nous pouvons nier la thèse de l'identité token-token, sans rejeter pour autant la thèse physicaliste puisqu'il semble tout à fait plausible que tous les événements soient composés d'événements physiques.

Considérons l'exemple paradigmatique de particuliers concrets que sont les objets. Je soutiens que deux objets numériquement distincts, par exemple une certaine masse de matière et vous-mêmes, peuvent coïncider dans une certaine région de l'espace et du temps. Les arguments en faveur de cette approche métaphysique insistent particulièrement sur le rejet de l'essentialisme méréologique<sup>25</sup>, et ils présupposent que chacun des deux objets nommés existent réellement comme des entités distinctes, malgré leur coïncidence dans l'espace et dans le temps<sup>26</sup>. Évidemment, l'existence de tels objets coïncidents porte à controverse, mais je considère très sérieusement la plausibilité d'arguments comme les suivants<sup>27</sup>. Considérons une petite statue, à un temps  $t$ , disons une réplique en argile du Discobole que vous avez placée sur une étagère, ainsi que la masse d'argile à partir de laquelle cette statue a été fabriquée. Au temps  $t$ , la statue et la masse d'argile ont toutes

---

21. Je ne suis pas un fan de la notion d'identité contingente, dont j'ai du mal à comprendre la cohérence.

22. Je parle de propriétés plutôt que de prédicats, tout comme le fait Davidson dans « Mental Events », bien que Davidson soit officiellement un nominaliste.

23. Voir Doepke, « Spatially Coinciding Objects » ; Thomson, « The Statue and the Clay » ; et Johnston, M., « Constitution is Not Identity », *Mind*, 99, 1998.

24. Les objets et les événements sont différents, mais je crois qu'ils partagent la propriété d'avoir des parties. Les événements n'existent pas dans leur totalité à des temps différents, alors que les objets se répètent à travers le temps.

25. Il y a deux versions importantes de cette thèse : un objet ne peut survivre (i) à la destruction d'une de ses parties, et (ii) s'il perd une de ses parties. Ces thèses semblent être fausses de la plupart des objets macroscopiques.

26. Peter Van Inwagen rejette l'existence de votre corps et celle d'artefacts ; voir Van Inwagen, P., *Material Beings*, Ithaca, Cornell University Press, 1990. Donc, selon lui, ces arguments sont faibles. Mais notons qu'un effort considérable d'argumentation non circulaire doit être déployé lorsqu'il s'agit de nier l'existence de notre propre corps.

27. Voir Doepke, « Spatially Coinciding Objects », et Thomson, « The Statue and the Clay ». L'exemple qui suit s'inspire d'exemples similaires de Thomson et de Baker, « Why Constitution Is Not Identity ».

leurs parties physiques en commun, mais il n'est pas nécessaire qu'elles aient toujours toutes leurs parties physiques en commun. La statue peut survivre à la perte de l'une de ses parties si, par exemple, lors d'une crise de dégoût esthétique, vous en brisez un des pieds que vous jetez ensuite par la fenêtre. Votre statue est endommagée mais elle est toujours sur l'étagère. Il n'en va pas de même pour la masse d'argile. Si celle-ci existe encore, ce qui paraît douteux, elle se sera dispersée : une de ses parties sera toujours sur l'étagère et le reste se retrouvera dans la cour. Briser l'un des pieds ne fait qu'endommager la statue, mais cela semble tout simplement détruire la masse d'argile ou, à tout le moins, la disperser dans l'espace<sup>28</sup>.

Certains philosophes ne sont pas convaincus par ces arguments en raison d'un certain scepticisme relativement au statut ontologique des artefacts<sup>29</sup>. Mais les arguments de ce genre semblent encore plus plausibles pour les « objets » humains :

- 1) Hermès peut perdre un pied et être toujours là où il était.
- 2) La masse de chair qui, à un certain temps *t*, coïncide dans l'espace et dans le temps avec Hermès ne peut perdre la partie qui coïncide avec le pied d'Hermès et être toujours là où elle était.

Donc : Hermès n'est pas la masse de chair.

Comme dans le cas de la statue, l'amputation du pied d'Hermès ne le détruit pas et ne le disperse pas. Mais si on considère la masse de chair dans son entier, alors ou bien l'amputation la disperse ou bien elle la détruit (et la masse de chair qui reste serait alors différente). Donc, Hermès n'est pas identique à la masse de chair. D'autres arguments de ce type peuvent être construits en utilisant des propriétés essentielles<sup>30</sup> :

- 1) Hermès est essentiellement une personne ou un être psychologique.
- 2) La masse de chair n'est pas essentiellement une personne ou un être psychologique.

Donc : Hermès n'est pas la masse de chair.

Plusieurs objections peuvent être adressées à ce type d'arguments, et leur défense exige beaucoup de travail. Mon but ici est simplement de développer ce point de vue métaphysique et d'en montrer la pertinence pour la discussion de l'argument de Kim. Si nous trouvons que ces thèses métaphysiques sont plausibles, et à mon avis elles le sont, alors le rejet du principe d'héritabilité causale devient tout à fait plausible.

28. Dans « The Statue and the Clay », Thomson suggère une autre version de cet exemple : vous remplacez la masse d'argile qui répugne à votre bon goût par une autre masse. Dans ce cas, la statue semble être complètement sur l'étagère, alors que la masse d'argile originale ne s'y trouve pas complètement.

29. Voir Van Inwagen, *Material Beings*.

30. Il s'agit ici d'une adaptation d'un argument de Baker, « Why Constitution Is Not Identity ».

La différence entre Hermès et la masse de chair, que l'on peut appeler son corps, est que Hermès est essentiellement une personne, tandis que la masse de chair pourrait survivre même si les conditions requises pour que Hermès soit une personne cessaient d'être satisfaites, par exemple la condition d'être une créature qui pense. Ainsi, les deux objets diffèrent relativement à une certaine propriété, à savoir la propriété d'être-essentiellement-une-créature-qui-pense, ou celle d'être-une-personne. Selon moi, il s'agit là d'une propriété causalement pertinente des objets et des événements qui l'exemplifient<sup>31</sup>. La propriété d'être-une-créature-qui-pense est causalement pertinente pour les objets qui pensent : ils peuvent raisonner et agir sur la base de leurs raisonnements. Pouvoir survivre sans penser est une différence causale qui distingue la masse de chair de la personne. (Il en va de même pour la capacité de survivre à la perte d'un tiers de votre masse corporelle.) Les différences modales entre des objets sont souvent des différences causales. Considérons les deux énoncés contrefactuels suivants, qui sont des énoncés causaux à propos de moi et de la masse de chair. Il semble y avoir des différences importantes entre eux. L'énoncé « Si je cessais de penser et que je m'effondrais sous la porte, je bloquerais le passage » est faux (à moins d'utiliser « je » de façon équivoque), car une personne ne peut survivre à la perte de la pensée et une personne qui n'existe pas ne bloque rien. Mais la masse de chair n'est pas assujettie à une telle limitation. En effet, il est plausible que l'énoncé suivant soit vrai : « Si la masse de chair cessait de penser et s'effondrait sous la porte, elle bloquerait le passage. » Ainsi certains objets distincts peuvent coïncider spatiotemporellement. Ils sont numériquement distincts, car différents ensembles d'énoncés contrefactuels s'y appliquent. Et il est plausible que ces énoncés contrefactuels révèlent des *différences et des propriétés causales*.

Illustrons cette idée en considérant le rôle causal/fonctionnel d'un objet. Le rôle causal/fonctionnel d'une propriété correspond aux généralisations nomologiques où elle figure. Pour les objets, nous pouvons spécifier le rôle causal en faisant la liste des divers énoncés contrefactuels qui s'y appliquent compte tenu de diverses circonstances actuelles et contrefactuelles. Ainsi le rôle causal de Hermès et celui de la masse de chair seront différents. Les propriétés qui rendent une personne différente de la masse de chair qui la compose (ou de son corps) sont des propriétés causalement efficaces. Être-une-créature-qui-pense ou être-essentiellement-une-créature-qui-pense sont causalement pertinentes.

Mais qu'est-ce que cela implique relativement au principe d'héritabilité causale? Notons d'abord que ce principe concerne la relation d'héritabilité de pouvoirs causaux, qui est une relation entre deux particuliers : soit

---

31. Cet argument peut sembler circulaire mais il ne l'est pas. Kim lui-même reconnaît l'existence de propriétés psychologiques causalement efficaces, à condition qu'elles soient identifiées à des propriétés physiques homogènes. Le point de vue que j'élabore ici ne dépend pas de la thèse de la réalisabilité multiple mais uniquement de l'efficacité causale des propriétés psychologiques.

l'occurrence d'une propriété mentale (être-une-créature-qui-pense) et celle d'une propriété physique (être-une-masse-de-chair). Mais si mes arguments sont corrects, alors il existe des différences causales entre ces deux particuliers, même s'ils coïncident dans l'espace et dans le temps. L'occurrence de la propriété être-une-personne, ou être-une-créature-qui-pense, ou même être-moi (si l'on est prêt à accepter une telle propriété), n'a pas nécessairement les mêmes pouvoirs causaux que ceux de l'occurrence coïncidente de la propriété physique. La relation d'héritabilité causale n'implique pas que ce soit nécessairement ou bien tous les pouvoirs causaux qui sont hérités, ou bien aucun. Ce qui est hérité *au temps t*, c'est, dans le meilleur des cas, les causes et les effets actuels : mais ceux-ci ne constituent pas la totalité de ce qui pourrait être hérité. De plus, les occurrences de propriétés survenantes ne sont pas des indigents : elles peuvent hériter mais elles travaillent également pour gagner leur vie, en ce sens qu'elles font aussi un travail explicatif puisqu'elles entrent dans des lois *ceteris paribus* ou dans des généralisations. Une occurrence de la propriété d'être-une-créature-qui-pense a des propriétés modales distinctes de celles d'une occurrence de la propriété d'être-une-masse-de-chair, même si cette dernière est constitutive de la première.

Mais que devrions-nous dire alors des propriétés de deuxième ordre? Kim soutient qu'il n'existe d'autres propriétés que les propriétés de premier ordre : « En quantifiant sur des propriétés, nous ne pouvons créer de nouvelles propriétés, pas plus que nous pouvons créer de nouveaux individus en quantifiant sur des individus<sup>32</sup> ». Considérons l'exemple suivant :

Être-dormitif est une propriété de deuxième ordre relativement à un ensemble de propriétés de base *B* (être-de-l'opium, être-un-tranquillisant-à-cheval, etc.) si et seulement si être-dormitif, c'est avoir une certaine propriété *P* (par exemple être-de-l'opium) telle que *D(P)* (par exemple, faire-dormir-un-sujet (opium)), où *D* spécifie une certaine condition sur les membres de *B*.

Mais qu'est-ce à dire au juste que l'opium satisfasse une certaine condition (causale)? Cela ressemble curieusement à une quantification : tous les membres de *B* possèdent la propriété de faire-dormir-un-sujet, ou la propriété de dormitivité! Les quantificateurs apparaissent deux fois plutôt qu'une dans cette définition des propriétés de deuxième ordre. Quantifier sur des propriétés n'amène pas de nouvelles propriétés à l'existence, mais quantifier sur deux ensembles de propriétés (les propriétés de base et celles qu'elles réalisent) nous engage ontologiquement à deux ensembles de propriétés. Bien sûr, le débat porte en grande partie sur cette question : y a-t-il deux ensembles de propriétés ou un seul? Si nous prenons aux sérieux les exemples que j'ai discutés (statue/argile et personne/corps), nous avons des raisons de croire qu'il y a bien deux ensembles de propriétés. Une occurrence d'opium constitue, ou compose, une occurrence de la dormitivité, mais celles-ci ne satisfont pas les mêmes contrefactuels causaux et ne sont donc pas identiques.

---

32. Kim, « The Mind-Body Problem : Taking Stock After Forty Years », p. 201.

Bien que je défende le point de vue selon lequel il existe des particuliers qui ne sont pas identiques aux particuliers qui les composent, dans la mesure où les premiers auraient des propriétés causales distinctes de celles des seconds, il semble assez évident que de tels particuliers et certaines de leurs propriétés ne peuvent exister qu'à condition que d'autres particuliers et d'autres propriétés existent. (C'est pourquoi ce point de vue n'est pas exactement cartésien.) Et c'est justement là que réside l'attrait d'une conception de la nature comme étant organisée dans une hiérarchie de niveaux, ainsi que la motivation à l'appui de la notion de propriété de deuxième ordre.

Cette problématique soulève plusieurs autres questions. D'où est-ce que les particuliers tiennent leurs pouvoirs causaux sinon complètement de leurs bases constitutives? Devrait-on croire que le domaine physique n'est pas causalement clos, que certains événements physiques n'ont pas de causes physiques? Je suis plutôt enclin à dire que les particuliers obtiennent leurs propriétés causales des propriétés qu'ils instancient. Que ces propriétés ne soient pas uniquement des propriétés physiques de premier ordre explique pourquoi les particuliers ne sont pas identiques aux particuliers qui les composent (la masse d'argile, de chair, d'atomes).

Mais qu'en est-il de la clôture causale du domaine physique, cette thèse selon laquelle tout changement physique a une explication physique complète, ou une cause physique complète? Si nous admettons que la statue soit une entité causale distincte de la masse d'argile qui la compose, n'admettons-nous pas du coup l'existence de pouvoirs causaux non physiques? Si nous admettons que les événements mentaux sont causalement distincts des événements physiques qui les composent, n'admettons-nous pas l'existence de pouvoirs causaux non physiques? Effectivement, il semble que j'admette l'existence de tels pouvoirs causaux, mais cela n'entraîne pas la conséquence additionnelle qu'il existerait des événements physiques qui ne sont pas causés par d'autres événements physiques. Autrement dit, je ne nie pas que chaque événement physique ait une cause physique complète et une explication physique complète, mais je soutiens simplement qu'il *existe* certains événements physiques qui sont causés par des événements constitués par des événements physiques mais non identiques à ceux-ci. Bien que le principe de clôture de la théorie physique soit correct, c'est-à-dire que tout événement physique a une cause physique, le principe de clôture *causale* du domaine physique doit être rejeté. En d'autres termes, même si chaque événement physique a une cause physique, certains de ces événements ont également une cause mentale : ils sont aussi causés par des événements mentaux distincts de leurs constituants physiques. Cette thèse peut paraître surprenante. Malgré les apparences, j'aimerais suggérer qu'elle ne se prononce pas tant sur la nature du domaine physique, mais qu'elle montre plutôt le caractère insaisissable de la nature des relations causales<sup>33</sup>.

---

33. Ce texte a été traduit de l'anglais par Paul Bernier. Je remercie Antonia LoLordo et Paul Bernier pour leurs commentaires portant sur des versions antérieures de ce texte.