

Max Kistler

Métaphysique de l'esprit et causalité chez Kim

préface à la traduction française de Jaegwon Kim, *Survenance et esprit*, Editions d'Ithaque, 2008, p. vii-xxviii.

Dans la communauté philosophique de langue anglaise, Jaegwon Kim, professeur à l'université Brown à Providence aux Etats-Unis, jouit d'un immense prestige. Dans le domaine de ce qu'on appelle la « philosophie de l'esprit », ce sont souvent ses thèses et arguments qui servent de points de départ des débats. La philosophie de l'esprit est cette branche de la philosophie qui prend pour objet le rapport entre notre esprit et le monde tel qu'il fait l'objet de la connaissance scientifique. Le travail de Kim dont le présent recueil contient un échantillon important, se distingue avant tout par sa rigueur exemplaire, par la clarté et transparence de son argumentation : Kim joue toujours « cartes sur table », expose clairement et explicitement la thèse qu'il entend démontrer, les prémisses sur lesquels il fonde son raisonnement et la structure de son argumentation. Ses arguments semblent souvent ne laisser aucune échappatoire : impossible de résister à une conclusion si l'on y parvient par une chaîne d'arguments dont chacun semble simple et valide. Or les conclusions de Kim sont parfois surprenantes ou même choquantes, pour être contraires non seulement au sens commun mais aussi à des doctrines philosophiques largement admises. Prenons la conception fonctionnaliste de l'esprit qui faisait, depuis les années 1960, l'objet d'un certain consensus dans toute la communauté des sciences cognitives, qui regroupent, autour de la philosophie de l'esprit, des branches de la psychologie, de la linguistique, de l'intelligence artificielle et des neurosciences. Selon le fonctionnalisme, le fait qu'une personne se trouve dans un état cognitif plutôt que dans un autre ne dépend pas de l'état précis de son cerveau, mais uniquement du réseau de causes et d'effets potentiels de l'état cognitif. Nous pouvons, vous et moi, ainsi que peut-être certains robots ou extra-terrestres de science-fiction, croire que l'herbe est verte, sans partager, ni entre nous ni avec les robots et extra-terrestres, aucun état cérébral précis qui donnerait lieu à cette croyance partagée. La doctrine fonctionnaliste communément admise voulait que des états cognitifs tels que les croyances et les désirs soient « multiréalisables » et « multiréalisés » : des systèmes cognitifs ayant des structures matérielles différentes peuvent néanmoins partager le même état cognitif. Avec une

terminologie empruntée à l'informatique, on dit que le même état cognitif peut être « réalisé », « implémenté » ou encore « incarné » dans différentes structures matérielles, comme on dit que le même logiciel peut être implémenté dans des ordinateurs ayant des structures physiques différentes. L'essence de l'état cognitif, son identité, est donc indépendante du substrat matériel ; elle est déterminée uniquement par le réseau de ses relations causales.

L'essai 16 du présent volume a fait date en ébranlant définitivement cette doctrine. Dans cet article, Kim montre que l'argument même qui a servi à motiver et justifier la conception fonctionnaliste de l'esprit, à savoir l'argument de la réalisabilité multiple, mène en réalité à la conclusion que les prétendus états cognitifs multiréalisés n'existent pas en tant que tels. Les seuls états réels car causalement efficaces sont les états réalisateurs, c'est-à-dire les états physiologiques (ou, peut-être, d'autres états physiques, dans le cas d'hypothétiques robots et extraterrestres doués de cognition). Cela a des conséquences assez dramatiques pour la conception que nous avons de nous-mêmes : nous sommes convaincus que ce sont nos décisions qui nous font agir et nos préférences et convictions qui mènent à nos décisions. Or la conclusion de l'argument de Kim est incompatible avec cette conviction : il montre en effet que les propriétés cognitives d'être une préférence, une conviction ou une décision n'ont aucune efficacité causale. Seules les propriétés neurophysiologiques des états et événements qui précèdent nos actions sont causalement efficaces dans la détermination de ces actions.

Le lecteur perplexe devant de tels résultats, et désireux de les contester, appréciera la limpidité du style argumentatif de Kim autant que celui qui se laisse convaincre. Les prémisses et l'argumentation étant clairement exposés, on sait comment il est possible de contester les conclusions : il faut argumenter contre une ou plusieurs prémisses. Je montrerai un peu plus loin de quelle manière on peut remettre en cause certaines de ces prémisses. Cependant, pour apprécier ce qui peut être contestable dans l'enchaînement apparemment sans faille des arguments kimien, il est indispensable de faire un petit détour par l'histoire toute récente de la philosophie.

Kim et l'histoire de la philosophie analytique : de l'empirisme logique à la métaphysique analytique

Kim appartient à une tradition philosophique qu'on qualifie souvent d'« analytique » et dont les origines remontent au cercle de Vienne des années 1920 et 1930. Pour les pères fondateurs de l'empirisme logique développé à Vienne et Berlin, tels que Carnap, Schlick ou

Neurath, la question de la réalité de l'esprit et de son statut par rapport au monde physique n'était à l'origine qu'un problème philosophique parmi d'autres : il s'agissait de faire progresser le débat philosophique à l'aide de l'analyse conceptuelle, au moyen d'outils de la logique formelle contemporaine. La stratégie de l'empirisme logique des philosophes viennois (et berlinois) était ce qu'on appelle aujourd'hui « déflationniste » : au lieu de prétendre à parvenir à une théorie métaphysique satisfaisante d'un domaine donné, on se contente d'analyser la logique de notre *discours* sur les phénomènes qui relèvent de ce domaine. Au lieu de découvrir si, par exemple, les nombres existent réellement, quelles sont leurs propriétés et quelles sont leurs relations avec les objets matériels, on se contente d'analyser la logique du *discours* arithmétique : au lieu de se demander quelles *propriétés* possèdent les nombres, on se demande quels *prédicats* il est légitime de leur attribuer. Les prédicats sont des mots ou d'autres expressions linguistiques, alors que les propriétés sont censées être ce qui leur correspond dans la réalité. On constate alors par exemple qu'il est correct de dire d'un nombre qu'il est premier ou qu'il est carré mais incorrect de dire qu'il est rapide ou bleu. Par ailleurs, comme l'exige la doctrine « vérificationniste » de la signification qui avait été développée au sein du cercle de Vienne, la signification d'un prédicat se fonde sur son association avec un critère opérationnel d'application. Pour notre exemple : un nombre naturel x est premier si et seulement si il n'existe aucun nombre naturel y autre que x et 1 tels que le résultat de la division de x par y est un nombre naturel.

Lorsque Carnap et d'autres appliquèrent cette stratégie de clarification philosophique par l'analyse logique à notre discours psychologique, ils parvinrent à une doctrine que les philosophes du cercle de Vienne appelaient « physicalisme »¹ - en un tout autre sens que celui associé aujourd'hui à ce terme² – et que l'on appelle aujourd'hui le béhaviorisme philosophique : dans une logique vérificationniste, l'attribution d'un prédicat psychologique à un individu a un sens bien déterminé à condition d'être équivalent à un énoncé observable. Un énoncé est observable si l'on peut déterminer sa valeur de vérité, au moins en principe, directement par l'observation. Ainsi, on peut parler d'une douleur de Paul quoique cette douleur ne soit pas elle-même directement observable par autrui, parce que l'énoncé « Paul a mal aux dents » est équivalent à toute une série d'énoncés qui sont observables au sens indiqué : « À la question : 'Qu'est-ce que tu as ?' Paul articule les mots : 'J'ai mal aux

¹ Voir p.ex. Rudolf Carnap, *Die physikalische Sprache als Universalsprache der Wissenschaft*, *Erkenntnis* 2 (1931), p. 432-465; Otto Neurath, *Physikalismus*, *Scientia* 50 (1931), p. 297-303 ; trad. *Physicalisme*, *ibid.* Suppl. p. 117-122.

² Aujourd'hui, et notamment dans les écrits de Kim, on désigne par « physicalisme » une doctrine ontologique. Voici une manière de l'exprimer : tous les objets sont physiques ou exclusivement composés d'objets physiques ; et toutes leurs propriétés sont physiques ou réductibles sans reste à des propriétés physiques.

dents' », ou « La pression sanguine de Paul, les processus de sa digestion, la rapidité de ses réactions présentent telles et telles altérations³ ».

Le béhaviorisme logique a été assez rapidement abandonné par la majorité des philosophes de tradition analytique⁴. Il est tout simplement faux que tout énoncé significatif sur l'état psychique d'un individu soit *équivalent* à un énoncé observable, où « équivalent » signifie qu'il lui est synonyme et que l'on peut donc découvrir cette équivalence de manière purement conceptuelle et *a priori*, sans recours à l'expérience. Il est impossible de découvrir le lien entre la douleur dentaire et l'augmentation de la pression sanguine sans recours à l'expérience. Pour des raisons analogues, la doctrine vérificationniste fut abandonnée de manière plus générale : il existe un grand nombre d'énoncés scientifiques – dont le caractère significatif ne fait pas de doute – qui non seulement ne portent pas directement sur des phénomènes observables mais qui ne sont pas non plus logiquement équivalents à des énoncés observables. Il est certes correct de dire qu'un énoncé de physique théorique, disons sur des photons ou sur la température de l'univers une seconde après le *big bang*, a des implications logiques sur des états de choses observables. Mais de telles implications présupposent un grand nombre d'autres prémisses empiriques. Cet épisode illustre bien ce qui fait la force de la tradition analytique en philosophie : une grande exigence en ce qui concerne la rigueur argumentative et l'absence de dogmatisme quant aux contenus. L'histoire de la philosophie analytique est marquée par de nombreux changements de doctrine, tel ou tel auteur abandonnant une thèse simplement parce qu'il a découvert un argument probant à son encontre.

Kim appartient à la seconde génération de la tradition analytique. Sa thèse de doctorat sur l'explication scientifique était dirigée par Carl G. Hempel qui fut, avec Carnap et Reichenbach, parmi les philosophes émigrés d'Autriche et d'Allemagne qui firent de la manière « analytique » de philosopher le style dominant aux Etats-Unis et par la suite dans de nombreux autres pays. Kim, parmi d'autres, montre qu'adopter un style où prime l'exposition limpide de la structure logique de l'argumentation est parfaitement compatible avec une grande ouverture d'esprit en ce qui concerne le contenu des thèses défendues et des prémisses

³ Ce sont les exemples que donne Carl G. Hempel, *L'analyse logique de la psychologie* (1935), in D. Fisette et P. Poirier (dir.), *Philosophie de l'esprit, vol. I : Psychologie du sens commun et sciences de l'esprit*. Paris, Vrin, 2002, p. 204.

⁴ Ainsi, Hempel n'a accepté de republier, en 1977, son article de 1935 (« L'analyse logique de la psychologie ») qu'accompagné d'une note qui précise que, « au moment où [l'article] paraissait en anglais [en 1949 ; M.K.], j'avais abandonné la forme étroite du physicalisme en termes de traduction » et que les mêmes « considérations qui m'ont amené à ces changements m'ont également conduit il y a longtemps à abandonner comme intenable la théorie vérificationniste de la 'signification empirique' d'un énoncé », C. G. Hempel, in Ned Block (éd.), *Readings in the Philosophy of Psychology, vol. I*, Cambridge (Mass.), Harvard University Press, 1980, p. 14.

acceptées. Ainsi, il figure parmi les auteurs responsables du renouveau de la métaphysique au sein même de la tradition qui avait accordé une importance fondamentale à l'abandon de la métaphysique. Il n'y a pourtant aucun paradoxe dans l'idée d'une « métaphysique analytique » : les doctrines positives, telles que le vérificationnisme, qui avait conduit les empiristes logiques à tenir la métaphysique pour impossible ou nécessairement confuse, ont été abandonnées. Par « analytique », on ne désigne ni un domaine de réflexion philosophique particulier ni une doctrine particulière, mais une méthode de travail fondée sur l'argumentation rigoureuse et sur l'expression explicite de la conclusion, des prémisses et de la structure logique des arguments, d'une part, et la référence à une certaine histoire argumentative, d'autre part. La métaphysique analytique, telle que la pratique Kim lorsqu'il s'interroge sur la nature de l'esprit et sur la question de savoir si nos propriétés cognitives peuvent être causalement efficaces, a abandonné la contrainte de l'empirisme logique de se limiter à l'analyse logique du langage ou du système conceptuel du sens commun. Kim et d'autres débattent de ces questions à nouveau comme de questions métaphysiques : ce qui donne sens à la question de savoir si nos états cognitifs existent et ont, par l'intermédiaire de nos actions, une influence causale sur le monde physique environnant, c'est que la recherche d'une réponse se soumet non seulement au verdict du sens commun et de la cohérence, mais aussi aux résultats et présupposés des sciences contemporaines.

La meilleure manière de montrer que ces contraintes « naturalistes » sont suffisantes pour soustraire la métaphysique à l'arbitraire auquel les empiristes logiques la croyaient condamnée, et lui permettent d'accomplir des progrès, consiste à en montrer des exemples. Dans l'article 10 (1982), Kim accepte, bien que de manière hypothétique, la conception fonctionnaliste des états mentaux comme intermédiaires causaux entre stimulations sensorielles et comportements. Elle implique notamment la thèse selon laquelle les « états internes postulés pour expliquer les relations entre entrées sensorielles et sorties comportementales *sont* des états psychologiques » (p. 192). Mais dans l'article 16 (1992), Kim parvient à la conclusion contraire que les intermédiaires causaux entre entrées sensorielles et sorties comportementales ne sont pas psychologiques mais exclusivement physiologiques. Ce ne sont pas des informations nouvelles, par exemple scientifiques, qui ont conduit Kim à changer d'avis ; au contraire, la nouvelle conclusion utilise les mêmes prémisses et notamment la thèse fonctionnaliste de la multiréalisabilité des états mentaux. Je reviendrai plus loin sur les raisons de ce changement.

Un autre changement important de doctrine – quoique ses conséquences soient un peu moins importantes pour les articles réunis ici – concerne l'interprétation de la survenance.

Dans une note à l'essai 11 (1985), Kim dit utiliser « le terme 'survenance' dans un sens [...] auquel ce qui survient est entièrement fixé lorsque la base de survenance est fixé », ce qui revient à donner au terme « survenance » un sens qui implique « la dépendance ou la détermination complète » (p. 213, note 28). En 1993 en revanche, Kim est parvenu à la conviction que le concept de survenance, tel qu'il l'a lui-même élaboré dans toute une série d'articles (réunis dans l'autre recueil⁵), n'a pas un tel sens : « Il me semble désormais que la conception de la survenance elle-même en tant qu'espèce particulière et distinctive de relation de dépendance » soit erronée ou du moins qu'elle induise en erreur, car « la survenance n'est pas une relation métaphysique 'profonde' ; plutôt, c'est une relation 'superficielle' qui [...] n'est pas en elle-même une théorie explicative de la relation corps-esprit ; plutôt, elle fait état des données qu'une telle théorie doit expliquer » (p. 167/8).

Deux principes méta-scientifiques et deux analyses de la causalité

Dans le reste de cette préface, je voudrais introduire à quelques arguments importants développés par Kim dans les articles du présent recueil. On s'intéressera surtout à deux principes méta-scientifiques qui leur servent de prémisses : le principe de la complétude du domaine physique et le principe d'exclusion explicative. Il apparaîtra que la plausibilité de ces principes dépend du modèle classique de l'explication causale de l'empirisme logique par rapport auquel la position de Kim n'est pas tout à fait tranchée.

Les articles 13, 14 et 17 développent un argument concernant la causalité mentale qui devait exercer une influence considérable sur le débat ultérieur. Kim s'attache à montrer que l'explication psychologique d'un événement ou état est incompatible avec son explication physique. Prenons l'événement physique de l'activation de la touche « e » par mon doigt sur le clavier d'ordinateur. À première vue, il semble envisageable d'expliquer (ou du moins de chercher à expliquer) cet événement de deux manières : sur le plan mental, en faisant référence à mon désir d'écrire le mot « et » dans le contexte d'une phrase dont je désire communiquer le contenu à mes lecteurs et à la croyance que je peux contribuer à satisfaire ce désir en appuyant sur la touche « e ». Dans la mesure où il s'agit d'un événement physique, il semble plausible qu'il en existe une autre explication de plus « bas niveau » : au niveau des événements physiques, l'activation de la touche s'explique par le mouvement de mon doigt qui s'explique à son tour par l'arrivée d'un signal nerveux aux muscles contrôlant mon bras et

⁵ Cf. **La survenance et l'esprit, Vol. 2, correspondant à la partie I de l'édition originale.** (NdÉ)

mes doigts. Ce signal est à son tour le dernier maillon de toute une chaîne d'événements neuronaux qui remonte au cortex moteur et au-delà. Aussi loin que l'on remonte cette chaîne d'événements « physiques » (plus précisément neurophysiologiques) le succès des sciences de la nature dans leur ensemble justifie la conviction que l'on ne sera jamais contraint de faire appel à des facteurs causaux non physiques. Il n'existe pas d'équivalent contemporain de la glande pinéale cartésienne où l'explication d'un événement physique nécessiterait le recours à une cause non physique. Kim considère l'existence, pour tout événement physique, d'une telle explication scientifique (ou physique – je reviendrai sur la différence) complète, comme un présupposé fondamental de la science dans son ensemble. Il s'agit d'un principe méta-scientifique qui conditionne la possibilité même de la science. Kim appelle ce principe le « principe de la clôture causale du domaine physique » : pour tout événement physique, à tout instant où cet événement possède une explication causale complète, il possède une explication causale complète purement physique.

L'existence parallèle de deux explications d'un même événement pose un problème philosophique formidable : l'une, psychologique, fait appel aux « raisons », notamment des croyances et désirs qui permettent de comprendre le mouvement de mes doigts comme une action rationnelle, et par là-même de la justifier, tandis que l'autre, physique, explique l'événement en mentionnant ses causes. Nous nous trouvons ici à un point stratégique pour la compréhension du rapport entre le corps et l'esprit. Il sera utile de comparer le raisonnement de Kim avec celui de Davidson qui fait l'objet d'une analyse minutieuse dans l'article 11. Davidson et Kim souscrivent tous deux à l'un des principes fondamentaux de l'empirisme logique qui remonte en réalité à l'origine de la philosophie moderne : que l'on peut analyser les notions de cause et d'explication causale à l'aide de la notion de loi ou d'explication nomologique⁶. Selon cette doctrine que Davidson appelle « le principe du caractère nomologique de la causalité⁷ », il existe pour chaque relation causale entre deux événements une ou plusieurs lois déterministes strictes⁸ de la nature, telles qu'il est possible de déduire

⁶ « Nomologique » dérive du mot grec *nomos* – la loi. Pour le débat sur l'interprétation de la notion de cause, voir M. Kistler, « La causalité dans la philosophie contemporaine », *Intellectica*, 38, 2004/1, p. 139-185 ; A. Barberousse, M. Kistler et P. Ludwig, *La philosophie des sciences au XXIe siècle*, Paris, Flammarion, Collection Champs-Université, 2000, chap. 4.

⁷ Cf. D. Davidson, « Mental Events », in D. Davidson, *Essays on Actions and Events*, Oxford, Clarendon Press, 1980, p. 3-19 ; trad. « Les événements mentaux » par P. Engel, *Actions et événements*, Paris, P.U.F., 1993, p. 277-302.

⁸ Le qualificatif « déterministe » s'oppose à « statistique » : une loi déterministe représente une variable comme fonction d'autres variables, alors que dans une loi statistique, c'est la *probabilité* des valeurs d'une variable qui est fonction d'autres variables. La loi selon laquelle la force de l'attraction gravitationnelle entre deux objets massifs est proportionnelle aux produits de leurs masses respectives est une loi déterministe ; la loi qui indique la probabilité avec laquelle, en fonction du temps, un atome d'une substance radioactive subit une décomposition

une description de l'événement effet à partir d'une description de l'événement cause et de l'énoncé de la loi pertinente (ou des lois pertinentes). C'est l'analyse « déductive-nomologique » de la causalité : dire que l'événement *a* est la cause de l'événement *b* signifie qu'il existe un argument déductif et nomologique (car faisant mention d'au moins une loi de la nature) qui contient parmi ses prémisses une description de *a* et dont la conclusion est une description de *b*.

Mais Davidson et Kim s'accordent aussi sur un autre point, pourtant extrêmement controversé : c'est la thèse, brillamment argumentée par Davidson⁹, selon laquelle l'explication d'une action qui consiste à mentionner les raisons qu'avait l'agent de l'effectuer, n'explique pleinement l'action que si elle montre également que l'agent a effectué l'action *pour ces raisons*. Davidson montre en effet que pour expliquer pourquoi l'acteur a agi comme il l'a fait, il ne suffit pas de montrer que l'agent entretient des croyances et des désirs qui ont avec une certaine description de l'action le lien *logique* qui lie une raison à ce dont elle constitue une raison (c'est-à-dire, le lien logique entre ce qui justifie et ce qui est justifié). L'agent pourrait avoir ces convictions et ces désirs et accomplir une action que ces convictions et ces désirs font apparaître comme rationnelle, tout en l'accomplissant *à cause* d'un autre mobile. En d'autres termes, on ne peut expliquer une action par les raisons *pour lesquelles* l'agent a agi qu'en montrant que ces raisons étaient aussi des causes de l'action : ce n'est pas le contenu des croyances et des désirs qui est alors pertinent, mais le support de leur représentation. Dans la mesure où les représentations des raisons sont réalisées dans des états, processus et événements cérébraux, elles peuvent être des causes.

Pourtant, le consensus entre Davidson et Kim s'arrête à un point crucial. Kim parvient à la conclusion que nos états mentaux manquent, *en tant que tels*, de pouvoirs causaux. Davidson se défend contre l'accusation de nier l'efficacité causale des propriétés mentales¹⁰,

spontanée, est une loi statistique. Une loi « stricte » est une loi qui n'admet aucune exception. Les lois des sciences « spéciales » sont généralement considérées comme non strictes. Sont appelées « spéciales » toutes les sciences en dehors de la physique : cette dernière s'applique à tout sans aucune restriction ; elle est donc la seule science générale et, partant, « non spéciale ». Cf. M. Kistler « Laws of Nature, Exceptions and Tropes », *Philosophia scientiae* 7 (2), 2003, p. 189-219, et « Lois, exceptions et dispositions », in Bruno Gnassounou et Max Kistler (éd.), *Les dispositions en philosophie et en sciences*, Paris, CNRS Editions, 2006, p. 175-194. La question est controversée de savoir s'il existe aussi des lois non strictes en physique, ou si toutes ou la plupart des lois physiques sont non strictes. Cf. N. Cartwright, *How the Laws of Physics Lie*, Oxford, Clarendon Press, 1983.

⁹ D. Davidson, *Actions, Reasons, and Causes* (1963), in D. Davidson, *Essays on Actions and Events*, Oxford, Clarendon Press, 1980, p. 3-19 ; trad. par P. Engel, *Actions et événements*, Paris, P.U.F., 1993, p. 15-36. Voir aussi M. Kistler, « Les causes des actions », *Le temps philosophique, Numéro spécial : L'action*, vol. 12 (2006), p. 141-175.

¹⁰ Voir « Thinking Causes », in J. Heil et A. Mele, *Mental Causation*, Oxford, Clarendon Press, 1993, p. 1-17. Brian McLaughlin formule l'accusation dans « On Davidson's Response to the Charge of Epiphenomenalism », *ibid.*, p. 27-40.

car il rejette la question de savoir « en vertu de quelle propriété » un événement en cause un autre. Davidson souscrit à l'autonomie et à l'irréductibilité de la psychologie et ainsi à une forme de dualisme. Bien entendu, il ne s'agit pas d'un dualisme ontologique postulant deux types de substances (ou de propriétés) mais seulement d'une dualité irréductible de formes d'explication. Selon Davidson, pour justifier la conviction intuitive que nous agissons pour des raisons, c'est-à-dire que les représentations des raisons sont des causes de nos actions, il suffit que des *événements* impliquant ces représentations soient parmi leurs causes ; il n'est pas nécessaire que ces événements soient causes *en tant que* représentations, c'est-à-dire en vertu de lois psychologiques. Il suffit que ces événements causent nos actions en tant qu'événements physiques, car, selon la doctrine davidsonienne du « monisme anomal », tous les événements mentaux sont aussi physiques ; autrement dit, tous les événements qui possèdent une description mentale possèdent aussi une description physique. De nombreux critiques ont objecté à cette analyse qu'elle ne rend pas pleinement justice à la conviction, pourtant justifiée précisément par Davidson dans son article de 1963, que nos raisons causent nos actions. Notre intuition ne dit pas seulement que, lorsque nous agissons pour des raisons, ces raisons sont des causes de notre action, mais aussi qu'elles ont causalement contribué à l'action *en tant que* raisons. Or il est difficile d'interpréter l'intuition selon laquelle les représentations des raisons conduisent à l'action en question en tant que raisons et donc en vertu de leur contenu, sans supposer qu'il existe des lois de niveau psychologique et intentionnel. L'existence de telles lois permettrait de justifier l'intuition selon laquelle les raisons peuvent être efficaces en tant que raisons : ces lois pourraient figurer dans l'explication causale (et déductive-nomologique) d'une action en plus de la condition initiale que l'agent s'est représenté les raisons en question. La difficulté vient alors du fait que l'hypothèse de l'existence de telles lois contredit la thèse davidsonienne de l'anomie du mental selon laquelle il n'existe pas de lois déterministes strictes portant sur les états psychologiques intentionnels.

Par un tout autre raisonnement, Kim parvient à une conclusion semblable : que les propriétés mentales ne sont pas efficaces en tant que telles. Contrairement à Davidson, Kim n'argumente pas directement contre l'existence de lois psychologiques. Son argument contre l'efficacité causale des propriétés mentales repose sur les deux principes mentionnés plus haut : le principe de complétude du domaine physique et le principe d'exclusion explicative.

Il existe une certaine tension entre les justifications que Kim propose, dans les articles de ce recueil, pour ces deux principes. En effet, ces justifications utilisent des conceptions différentes, et peut-être incompatibles, de la causalité : la justification du principe de la

clôture causale du monde physique présuppose la conception déductive-nomologique, alors que la justification du principe d'exclusion présuppose l'analyse contrefactuelle de la causalité.

Selon le principe de complétude du domaine physique, « tout événement physique qui a une cause à l'instant t a une cause physique à l'instant t ». (p. 280) Ce principe a des conséquences considérables : joint au principe d'exclusion explicative¹¹, il conduit à remettre en cause non seulement les explications psychologiques mais toutes les explications causales non physiques¹². Kim justifie le principe de clôture par l'autarcie de la physique en ce qui concerne l'explication (déductive-nomologique) des événements physiques. Ce principe a le statut d'une hypothèse méta-scientifique qui est confirmée par induction à partir des succès de la physique : « si la clôture causale faisait défaut, notre physique aurait besoin de faire référence, de manière essentielle, à des agents non physiques, peut-être des âmes cartésiennes et leurs propriétés psychiques » (p. 280). Kim s'appuie ici implicitement sur le principe du caractère nomologique de la causalité¹³ : son raisonnement présuppose que le fait qu'un événement P en cause un autre P^* est équivalent au fait – ou du moins implique le fait qu'il existe une explication déductive-nomologique de P^* en vertu de P et de lois exprimant une dépendance entre P et P^* . Dans la phrase citée, Kim utilise la contraposée de cette implication : s'il apparaît que, pour un événement P^* donné, il n'y a pas d'explication déductive-nomologique (DN) à l'intérieur de la physique, cela signifie que P^* n'a pas de cause (purement) physique.

Mais l'évaluation de la thèse de clôture est loin d'être simple. Sa plausibilité dépend de deux choses : d'une part du choix de considérer la causalité comme une relation entre événements *particuliers* ou entre *faits*. On peut chercher la cause d'un *événement concret* ayant de nombreuses propriétés ou on peut chercher la cause *du fait que l'événement a une certaine propriété*¹⁴. On peut chercher la cause de *l'événement* de l'éloignement de la boule de billard après la collision, ou rechercher une explication causale du *fait* que la boule de billard a une vitesse d'1 cm/s. La première manière de concevoir les causes et les effets est

¹¹ Dans sa version faible selon laquelle il n'y a jamais deux explications causales complètes et indépendantes d'un événement donné. Je reviens plus loin sur la distinction entre les versions faible et forte du principe d'exclusion.

¹² Je reviens sur le problème de la « généralisation » du problème de la causalité mentale dans un instant.

¹³ De la même manière, dans l'argument contre l'efficacité du mental du chap. 17, Kim passe sans justification explicite de l'hypothèse que P est la cause de P^* à l'hypothèse selon laquelle « il existe une loi appropriée qui lie les instances de P aux instances de P^* » (p. 354). Autrement dit, il présuppose le principe du caractère nomologique de la causalité.

¹⁴ Pour la distinction entre causalité entre événements et responsabilité causale entre faits, voir M. Kistler, *Causalité et lois de la nature*, et M. Kistler, « La causalité comme transfert et dépendance nomique », *Philosophie* 89 (2006), p. 53-77.

celle de Davidson qui considère les événements comme des événements concrets, la seconde est celle de Kim. Kim, lui aussi, appelle « événements » les termes des relations causales, mais les « événements » kimiens sont en réalité ce que d'autres auteurs appellent des faits. Un tel fait consiste en la possession d'une propriété par un particulier. En effet, la plupart des explications causales vise l'explication d'un *fait* : on ne cherche pas à expliquer l'événement concret dans toute sa complexité, mais on cherche l'explication causale d'un seul fait le concernant, comme, dans notre exemple, le fait que la boule ait la vitesse 1 cm/s.

Dans la perspective de l'analyse de la causalité selon laquelle toute relation causale entre deux événements repose sur la transmission d'une quantité déterminée d'énergie (ou d'une autre grandeur conservée)¹⁵, la clôture causale devient triviale : un événement donneur ou receveur d'énergie est nécessairement physique. Mais elle est loin d'être triviale lorsqu'on considère la causalité comme la relation qui fait l'objet des explications causales : s'il s'agit d'expliquer un *fait* au sens d'en produire une déduction nomologique, c'est une question empirique ouverte de savoir si pour tous les événements physiques, il existe une telle explication qui ne fait appel, dans ses prémisses, qu'aux faits et lois physiques. L'évaluation de cette dernière thèse dépend d'une seconde prise de position, importante mais quelque peu occultée dans la réflexion de Kim : on peut prendre le mot « physique » en un sens étroit ou large. Au sens étroit, sont physiques les prédicats qui sont effectivement utilisés dans la science physique et les lois qui font l'objet d'hypothèses explicitement exprimées en physique. Au sens large, est physique tout ce qui n'est pas mental ou social, c'est-à-dire l'ensemble des prédicats et lois qui font l'objet de toutes les sciences dites exactes. La thèse de la complétude paraît plus plausible si l'on prend le mot « physique » au sens large que si on le prend au sens étroit : il est impossible, de fait et pratiquement, d'expliquer le développement d'un organisme vivant dans des termes physiques au sens étroit. Le défenseur du principe de clôture peut répliquer de deux manières. Soit il répond que cette impossibilité n'est que pratique et provisoire et que rien ne s'y oppose *en principe*, soit il admet que c'est impossible et se rabat sur l'interprétation large du prédicat « physique » : l'explication du développement de l'organisme est possible en faisant appel à des conditions initiales et des lois exprimées dans un vocabulaire biologique et chimique. La première réponse affaiblit la force de la justification inductive de la thèse : on ne peut pas prétendre que la physique, telle qu'elle existe effectivement aujourd'hui, puisse expliquer tous les événements physiques à

¹⁵ Cette conception est défendue dans M. Kistler « Reducing Causality to Transmission », *Erkenntnis* 48 (1998), pp. 1-24, et M. Kistler, *Causalité et lois de la nature*. Voir aussi A. Barberousse, M. Kistler et P. Ludwig, *La philosophie des sciences au XXe siècle*, chap. 4 et M. Kistler, « La causalité dans la philosophie contemporaine ».

partir de seules prémisses physiques. Une telle explication est pratiquement impossible, non seulement pour la plupart des événements décrits en vocabulaire chimique et biologique mais aussi pour des événements macroscopiques identifiés avec des concepts du sens commun, tels que des accidents de voiture. L'induction justifie donc plutôt la thèse contraire selon laquelle, en général, il *n'est pas* possible d'expliquer un événement quelconque (non mental ou social) en termes physiques au sens étroit. La seconde réponse pose le problème de la justification du présupposé d'une différence fondamentale entre les sciences « physiques » (au sens large) mais non physiques au sens étroit, telles que la chimie, la géologie, la biologie ou les neurosciences d'un côté, et la psychologie et la sociologie de l'autre côté. En effet, les raisons qui obligent le scientifique à avoir par exemple recours, outre aux concepts et lois physiques, à des concepts et lois neuroscientifiques pour expliquer un événement cérébral, semblent être du même ordre que celles qui l'obligent à avoir recours à des concepts et à des hypothèses psychologiques pour expliquer le comportement d'animaux ou d'humains.

La justification du principe d'exclusion causale requiert une autre conception de la causalité que celle qui la réduit à une explication déductive-nomologique (DN). Sur le plan de l'explication DN, rien ne s'oppose à la co-existence d'un nombre arbitrairement grand d'explications du même phénomène à partir de prémisses différentes et même indépendantes les unes des autres. Dans la mesure où l'on analyse la relation causale entre l'événement *a* et l'événement *b* en termes de l'existence d'une explication DN dont la conclusion est une description de *b* et dont les prémisses contiennent une description de *a* comme partie essentielle, rien ne semble s'opposer à la coexistence de plusieurs causes complètes a_1, a_2, \dots , du même événement *b*. En effet, Kim justifie le principe d'exclusion causale dans le cadre d'une autre analyse de la causalité : l'analyse contrefactuelle (essai 13, p. 253).

Il s'agit de deux conceptions de la causalité radicalement différentes : l'analyse DN conçoit la causalité sur un plan linguistique et argumentatif. On conçoit la cause, l'effet et le lien qui les unit, sur le plan du langage : en parlant de la cause et de l'effet, on entend parler de *descriptions* de la cause et de l'effet ; par lien causal, on entend un lien explicatif et argumentatif analysé en termes d'arguments valides, ayant la description de la cause et des énoncés de lois de la nature comme prémisses, et la description de l'effet comme conclusion.

En revanche, lorsqu'on demande, dans le cadre de l'analyse contrefactuelle, si l'effet se serait produit si l'une ou l'autre de deux causes potentielles *a* et a^* ne s'était pas produite, on parle d'événements particuliers, et non de leurs descriptions. On considère que la valeur de vérité de tels contrefactuels est déterminée par certains mondes possibles différents de notre monde actuel ; ce qui nous permet d'évaluer un contrefactuel, c'est la possibilité d'intervenir

et notamment de faire des expériences. Nous ne pouvons pas observer ce qui se passe dans d'autres mondes possibles mais nous pouvons décider de produire des événements. Pour évaluer le contrefactuel : « si a se produisait mais non a^* , b ne se produirait pas », nous pouvons créer une situation expérimentale où un événement du type de a se produit, mais aucun événement du type de a^* . Puis nous observons si un événement du type de b se produit.

Dans le cadre de la causalité entre événements davidsoniens, l'exclusion explicative ne serait pas plausible : rien n'empêche la même séquence d'événements de faire l'objet de deux explications causales indépendantes. Le fait que la boule A ait frappé la boule B, au repos, avec une vitesse d'1 cm/s explique le fait que la boule B s'éloigne, après le choc, à cette même vitesse (si l'on suppose que les masses de A et B sont égales et que le choc est élastique), mais le fait que la boule A ait frappé la boule B alors qu'elle était couverte d'une peinture rouge qui déteint explique le fait que la boule B porte une marque rouge. Ce n'est que lorsque l'on cherche les causes des faits, non des événements, qu'il est plausible de dire « que l'on ne peut pas permettre ... qu'il arrive que [nos explications] racontent deux histoires différentes de la même région de la carte causale » (p. 255). Par ailleurs, comme nous l'avons vu, sur le plan de la causalité entre faits, l'exclusion explicative ne se justifie que dans le cadre de l'analyse contrefactuelle mais non dans le cadre de l'analyse DN.

Le fait que Kim utilise deux conceptions de la causalité – respectivement, déductive-nomologique et contrefactuelle – dans la justification des deux principes sur lesquels repose son argumentation, pose problème dans la mesure où il est généralement admis que ces analyses sont incompatibles¹⁶.

Les versions faible et forte du principe d'exclusion

Nous avons vu que la position de Kim par rapport à la doctrine traditionnelle de l'équivalence entre explication scientifique, explication causale et explication déductive-

¹⁶ Dans un article récent (J. Kim, *Causation and Mental Causation*, in : Brian McLaughlin and Jonathan Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, Oxford, Blackwell, 2007), souscrit à l'analyse de la causalité en tant que « causalité productrice » selon laquelle la connexion causale est fondée sur la transmission d'énergie, de quantité de mouvement ou d'une autre grandeur conservée. (J'ai développé cette analyse dans « Reducing Causality to Transmission » et dans *Causalité et lois de la nature*, Paris, Vrin, 1999.) Il soutient en particulier que le concept de causalité productrice est le seul qui correspond à l'intuition de causalité mentale à l'œuvre dans nos jugements sur l'origine causale des actions. Pour être à l'origine causale des conséquences de mes actions, je, ou l'une de mes décisions, doit être la cause productrice des mouvements corporels adéquats. En revanche, Kim explique que la dépendance contrefactuelle, tout autant que la dépendance nomologique, ne sont pas suffisantes pour l'existence d'une relation causale : ces relations peuvent exister entre des événements qui ne sont pas cause l'une de l'autre s'ils sont deux effets d'une cause commune, ou s'ils font partie d'une suite d'épiphénomènes accompagnant un processus causal réel.

nomologique ne semble pas, dans les articles du présent recueil, tout à fait tranchée. Kim la présuppose souvent, pour s'en démarquer à d'autres moments. Je voudrais mentionner une difficulté dans l'interprétation du principe d'exclusion qui semble elle aussi être due à cette hésitation, et qui menace la validité de l'argument de Kim contre l'efficacité causale du mental. Kim semble hésiter entre l'affirmation que le principe d'exclusion vaut pour *toutes* les explications scientifiques et la thèse plus prudente selon laquelle il ne vaut que pour les exclusions *causales*. Notons que cette distinction n'a de sens que si l'on abandonne la doctrine traditionnelle. « J'avancerai des arguments [en faveur de la thèse de l'exclusion explicative] surtout pour les explications causales, avançant en même temps certaines considérations générales qui, je l'espère, feront apparaître le principe d'exclusion comme une contrainte plausible sur les explications en général » (p. 239). Kim semble dire qu'il est convaincu du fait que le principe vaut en général, mais estime n'avoir d'arguments concluants que pour la version plus faible selon laquelle il ne vaut que pour les explications causales. Cependant, il contredit cette déclaration lorsqu'il donne lui-même une raison de penser que le principe d'exclusion ne vaut pas au-delà des explications causales : dans une note, il suggère que le principe d'exclusion *ne s'applique pas* lorsqu'il s'agit d'expliquer une action. Kim suggère qu'il peut y avoir deux explications complètes et indépendantes d'une même action si l'une d'elles est non causale. Lorsqu'on est confronté à deux explications d'une même action, dit Kim, de sorte que le principe d'exclusion met en doute la complétude ou l'indépendance de l'une d'elles, « une manière dont on pourrait essayer d'éliminer l'incompatibilité est d'interpréter [l'une des deux explications] comme relevant d'un mode de comprendre les actions qui est fondamentalement *non causal* » (p. 240, note 4).

La différence entre la version forte et faible du principe d'exclusion s'avère cruciale dans l'argument que Kim développe au chapitre 17, pour montrer que la causalité mentale est conceptuellement impossible¹⁷. Notre esprit n'a pas de pouvoirs causaux propres : il n'est jamais littéralement vrai qu'une propriété mentale est causalement efficace. Le fait d'effectuer, par exemple, un raisonnement, ne peut avoir d'influence causale sur la décision que nous prenons à sa suite, que dans la mesure où le fait d'effectuer ce raisonnement est *identique* au fait de faire l'objet d'un processus neurophysiologique. Le cadre métaphysique dans lequel Kim développe son argument est le suivant : on part de l'hypothèse de l'identité des événements, processus et états mentaux avec des événements, processus et états neurophysiologiques sous-jacents, *en tant qu'entités particulières*. Par « particulier », on

¹⁷ Kim a développé cet argument dans son ouvrage *Mind in a Physical World*, MIT Press 1998 ; trad. *L'esprit dans un monde physique*, Paris, Syllepse, 2006.

entend quelque chose qui possède, à l'instar d'une substance traditionnelle, de nombreuses propriétés. Une personne est par exemple un particulier qui possède à la fois des propriétés physiques (telles que la propriété de peser 70 kg) et mentales (telles que la propriété de croire que 5 est un nombre premier). La question de la causalité mentale est celle de savoir si les propriétés mentales sont causalement efficaces¹⁸ : est-il possible qu'un événement qui implique une personne provoque un changement *en vertu de ses propriétés mentales* ? L'argument contre la causalité mentale procède en deux étapes : dans la première, Kim considère la possibilité qu'une propriété mentale soit causalement efficace au sens de causer une autre propriété mentale – comme lorsque la dernière étape d'une chaîne de raisonnement cause, *en vertu de la propriété mentale* d'être une représentation des raisons d'agir, la prise d'une décision en tant qu'instance de la propriété mentale de prendre la décision d'une action déterminée. Il montre qu'une telle hypothétique causalité mentale présuppose une forme de causalité « descendante » (*downward causation*) où une propriété mentale cause une propriété physique. Si une propriété mentale M_1 pouvait causer une propriété mentale M_2 , elle ne pourrait le faire qu'indirectement, en causant directement la propriété physique P_2 , sous-jacente à M_2 . Je reviens toute de suite à son argument pour cette thèse. Dans la seconde étape de l'argument, il montre, à partir des principes de complétude du domaine physique et du principe d'exclusion des explications causales qu'il ne peut pas y avoir de telle causalité descendante.

Je voudrais attirer l'attention sur le fait que la première étape de l'argument de Kim présuppose le principe d'exclusion *dans sa version forte, plus générale*. En effet, Kim raisonne ainsi : la thèse de la survenance locale forte garantit qu'il existe, pour toute propriété mentale, une explication DN qui explique cette propriété à partir d'une propriété physique ou neurophysiologique sous-jacente. En ce qui concerne la propriété mentale M_2 , par exemple d'être une décision d'écrire le mot « mental », la survenance forte locale garantit l'existence d'un ensemble de types de propriétés physiques (ou physiologiques) $\{P_{21}, \dots, P_{2n}\}$ telles que nécessairement, si un objet ou un événement a une de ces propriétés P_{2i} , il a aussi la propriété mentale M_2 d'être une décision d'écrire le mot « mental ». La survenance locale forte garantit donc qu'il y a nécessairement une explication non causale du fait qu'un objet ou événement

¹⁸ Lorsque Kim pose la question de savoir si une propriété mentale peut causer d'autres propriétés, il s'agit d'« instances de propriétés » : les propriétés *universelles* ne peuvent pas être des termes de relations causales. Si une brique tombe sur mon pied, ce n'est pas la propriété universelle d'avoir une masse de 5 kg qui cause la fracture de mon orteil, mais l'instance de cette propriété dans cette brique à cet instant. (L'exemple est de Keith Campbell, *Abstract Particulars*, Oxford, Blackwell, 1990, p. 113). On pourrait aussi poser la question de l'efficacité causale du mental en demandant si le *fait* que la personne possède la propriété mentale M_1 peut être causalement responsable du *fait* qu'elle possède à l'instant suivant la propriété mentale M_2 .

possède la propriété mentale M_2 , à partir du fait qu'il possède la propriété physique sous-jacente P_{2i} . À ce point, l'argument de Kim dépend du principe d'exclusion dans sa version forte, générale : « Je crois, dit-il, que ces deux explications de la présence de M_2 à cette occasion [à savoir : (1) par la cause mentale M_1 et (2) par la propriété neurophysiologique sous-jacente P_{2i} ; M.K.] créent une tension et doivent être réconciliées » (p. 351). Dans une note, il justifie cette affirmation par un renvoi au chapitre 13 dans lequel, nous l'avons vu, il n'argumente que pour la version *faible*, et plus étroite, du principe d'exclusion : « Cela est, dit-il, essentiellement identique à la situation dans laquelle nous nous trouvons lorsqu'on nous donne deux causes indépendantes d'un même événement dont chacune prétend être une cause suffisante » (p. 351, note 30). Or la situation ne serait « essentiellement identique » que s'il n'y avait aucune différence entre une explication scientifique et une explication causale, autrement dit, si la doctrine traditionnelle de l'empirisme logique selon laquelle toutes les explications scientifiques sont causales était justifiée. Dès lors que l'on abandonne cette doctrine – comme Kim le fait lui-même – nous n'avons plus de raisons de souscrire au principe d'exclusion dans sa version forte. Or, le principe faible d'exclusion ne nous donne aucune raison de penser qu'il y a une « tension » entre l'explication de la prise de décision M_2 (1) par la cause mentale M_1 et (2) par la propriété physiologique P_{2i} sous-jacente à M_2 . L'explication de M_2 par M_1 est causale, alors que l'explication de M_2 par P_{2i} est non causale.

Le problème posé par l'hésitation entre les deux versions du principe d'exclusion¹⁹ apparaît comme un indice de l'émancipation progressive et partielle de Kim par rapport à la doctrine de l'empirisme logique selon laquelle toute explication scientifique est causale. On ne peut en effet envisager la différence entre les deux versions sans avoir abandonné cette doctrine. Kim a lui-même contribué à son abandon, notamment en analysant la logique de différentes formes d'explication non causale²⁰. Étant donné qu'il existe des explications non causales, il est clair pour lui que les deux versions ne sont pas équivalentes. Cependant, il lui arrive, et cela même à des moments stratégiques, d'oublier l'existence d'explications non causales, autrement dit de retourner à la doctrine traditionnelle. Ici, cela le conduit à négliger la différence entre les deux versions du principe d'exclusion.

Les espèces naturelles des sciences spéciales et l'irréalité des espèces psychologiques

Je voudrais, pour finir, mentionner un dernier point stratégique où l'argumentation de Kim dépend d'une prise de position par rapport à la doctrine traditionnelle de l'empirisme logique. Dans les essais 14 et 16, Kim montre que la thèse de la multiréalisabilité des espèces naturelles psychologiques implique que ces espèces ne correspondent pas à des pouvoirs

¹⁹ Dans des écrits ultérieurs, Kim a clarifié sa position : il distingue désormais explicitement entre le « principe d'exclusion causale » (qui correspond à ce que j'ai appelé la version faible du principe d'exclusion) et le « principe d'exclusion déterminative/générative » (qui correspond à ce que j'ai appelé la version forte). Cf. J. Kim, *Mental Causation and Consciousness : The Two Mind-Body Problems for the Physicalist*, in : C. Gillett et B. Loewer (éds.), *Physicalism and its Discontents*, Cambridge, Cambridge University Press, 2001, p. 271-283, à p. 276/7 ; et J. Kim, *Physicalism, or Something Near Enough*, Princeton, Princeton University Press, 2005, p. 17). Dans l'article « The Two Mind-Body Problems... », il apparaît clairement que son argument selon lequel la causalité mentale n'est possible que si elle passe par la causalité « descendante », requiert la version forte du principe d'exclusion. Mais son seul argument en faveur du principe fort consiste à l'assimiler au principe faible : il dit en effet que le principe d'exclusion garde toute sa plausibilité si on y substitue la détermination en général à la détermination causale. « La raison fondamentale de souscrire au principe plus large [c'est-à-dire plus fort ; M.K.] est essentiellement la même » (p. 277) que celle de souscrire au principe faible (ou étroit). Dans *Physicalism, or Something Near Enough*, Kim dit préférer « ne faire appel à aucun principe général » (p. 41, note 8) à ce stade de l'argument. En réalité, l'argument dépend toujours de la partie controversée du principe d'exclusion dans sa version forte, selon laquelle une instance de propriété ne peut pas être à la fois expliquée par une explication causale et une explication non-causale. Kim appelle cette thèse le « lemme d'Edwards » : « Il y a une tension entre détermination 'verticale' et causalité 'horizontale'. En fait, la détermination verticale exclut la causalité horizontale » (p. 36). Dans la doctrine du théologien et philosophe Jonathan Edwards (*Doctrines of Original Sin Defended*, 1758), ce principe d'exclusion est utilisé pour justifier une doctrine d'inspiration occasionnaliste : les objets ne persistent pas dans le temps – il n'y a pas de relations causales « horizontales » entre leurs parties temporelles – parce que, à chaque instant, Dieu cause leur existence de manière « verticale », et parce que cette détermination « génératrice » exclut toute détermination causale. On peut répondre que ce résultat conduit plutôt à une réduction à l'absurde du lemme d'Edwards : dans la mesure où il mène à la conclusion absurde que les seules relations causales réelles sont celles qui ont leur origine dans Dieu, le lemme d'Edwards n'est pas plausible.

²⁰ Voir « Connexions non causales (*Noncausal Connections*) », Essai 2 dans Kim, *La survenance et l'esprit*, Vol. 2, correspondant à la partie I de l'édition originale.

causaux propres : une instance d'une espèce psychologique peut « faire une différence » causale, autrement dit être causalement efficace, seulement dans la mesure où elle est identique à une instance d'espèce physique (ou physiologique). Elle ne peut donc provoquer aucun effet *en tant qu'*instance d'espèce mentale, par exemple en tant que prise de décision. En d'autres termes, l'espèce mentale en tant que telle n'a aucun pouvoir causal.

L'argument avec lequel Kim réfute l'existence des pouvoirs causaux des espèces mentales ne s'appuie pas sur l'analyse traditionnelle de la causalité en termes d'explication DN. Sa démarche se distingue à cet égard de celle de Davidson qui s'appuie sur la thèse selon laquelle il n'existe pas de lois psychologiques déterministes strictes. Selon Davidson, l'explication causale requiert de telles lois. Toute explication causale, même lorsqu'il s'agit de propriétés ou de faits mentaux, ne peut donc s'appuyer que sur des lois physiques. Kim ne fait pas appel à l'inexistence de lois psychologiques déterministes et strictes. Cela s'explique en partie par une différence fondamentale dans ce que l'un et l'autre entend par « psychologique » : le but de Davidson est de remettre en cause la pertinence causale des espèces naturelles *intentionnelles*, telles que les croyances et autres attitudes propositionnelles. En revanche, l'argument de Kim a une visée beaucoup plus large : l'argument de Kim remet en cause l'existence de pouvoirs propres de toutes les espèces psychologiques, y compris de celles qu'identifie la psychologie scientifique, telles que, par exemple, la force d'association entre deux stimuli dans la théorie du conditionnement classique. Les arguments davidsoniens contre l'impossibilité de lois entre états intentionnels reposent sur le fait que l'attribution de tels états est soumise à des normes de rationalité. Ces arguments ne peuvent pas être transposés pour remettre en cause l'existence de lois portant sur des états postulés par la psychologie scientifique, sur la base de critères observationnels. Rien ne semble distinguer les lois découvertes dans le domaine du conditionnement classique des lois d'autres sciences exactes non physiques. Si l'on voulait contester qu'il s'agisse de lois sous prétexte qu'elles ne valent que *ceteris paribus* (toutes choses étant égales par ailleurs), on conteste du même coup l'existence de lois dans toutes les sciences en dehors de la physique fondamentale. Kim ne peut donc pas faire appel à l'analyse DN pour contester l'efficacité causale du mental : l'analyse DN tend plutôt à accréditer la thèse contraire selon laquelle la psychologie identifie bel et bien des propriétés efficaces. Étant donné que, par exemple, la force d'association entre stimuli obéit à des lois qui l'associent à des caractéristiques mesurables du protocole expérimental, l'analyse DN de la causalité conduit à la conclusion que cette force est causalement efficace : elle peut figurer dans une explication causale du comportement d'un animal ou d'une personne ayant subi un conditionnement.

L'argument de Kim en faveur de l'irréalité des propriétés psychologiques, ou dans ses propres termes, de la conclusion que « nos termes et concepts mentaux généraux [par opposition aux termes et concepts mentaux qui sont spécifiques à une espèce biologique, M.K.] ne font pas référence à des propriétés et espèces dans le monde » (p. 335), passe par la thèse de la multiréalisabilité des propriétés et espèces psychologiques. Cela rend l'argument vulnérable aux objections contre la thèse de la multiréalisabilité des espèces psychologiques. Des auteurs comme Bechtel et Mundale²¹ font en effet valoir que la recherche en neurosciences cognitives repose sur la présupposition que les capacités cognitives dont on cherche à identifier le mécanisme physiologique sous-jacent *ne sont pas* multiréalisées : en faisant des expériences sur des macaques ou des singes hibou pour comprendre des aspects fondamentaux du processus de perception visuelle humaine, on suppose que ces espèces partagent entre elles et avec l'espèce humaine, la structure fondamentale du système visuel à la fois sur le plan cognitif et sur le plan neurophysiologique.

Cependant, étant donné que la réalité et l'étendue de la réalisation multiple parmi les propriétés psychologiques est une question empirique, nous pouvons la laisser de côté ici. Admettons donc avec Kim – et avec le dogme fonctionnaliste contre lequel il argumente – que les propriétés psychologiques sont vraiment multiréalisables. Si l'argument avancé par Kim contre la thèse de l'autonomie et de l'irréductibilité de la psychologie et donc contre la réalité des propriétés psychologiques, a une si grande force dialectique, c'est parce qu'il s'appuie sur les prémisses mêmes qu'utilisent les défenseurs de cette thèse : selon la doctrine du fonctionnalisme « anti-réductionniste », les propriétés psychologiques sont irréductibles parce que les espèces neurophysiologiques (ou les espèces appartenant à d'autres sciences de plus bas niveau que la psychologie) qui leur sont sous-jacentes, sont hétérogènes. On ne pourrait, disent les anti-réductionnistes, établir un pont conceptuel entre le niveau psychologique et le niveau neurophysiologique qu'à condition d'associer une espèce mentale à une *disjonction* d'espèces neurophysiologiques hétérogènes. Les anti-réductionnistes font valoir qu'une disjonction hétérogène n'est pas une espèce naturelle, et qu'elle ne peut donc figurer dans aucune loi de la nature : elle ne peut en particulier pas figurer dans un « principe de correspondance » requis, dans le modèle de la réduction classique de Nagel²², pour réduire la propriété psychologique. À cela, Kim réplique que, étant donné que les pouvoirs causaux de

²¹ Bechtel, William, et Mundale, Jennifer, Multiple Realizability Revisited: Linking Cognitive and Neural States, *Philosophy of Science* 66 (1999), p. 175-207.

²² E. Nagel, *The Structure of Science*, London, Routledge and Kegan Paul, 1961. Voir M. Kistler, « Matérialisme et réduction de l'esprit », dans Jean Dubessy, Guillaume Lécointre et Marc Silberstein (dir.), *Les matérialismes (et leurs détracteurs)*, Paris, Syllepse, 2004, p. 309-339.

toute instance de propriété psychologique sont les pouvoirs causaux de l'instance physique sous-jacente²³, la propriété psychologique n'est pas naturelle non plus : ce qui est réel en elle, ce sont ces pouvoirs causaux, et les différents pouvoirs causaux des différentes instances de la propriété psychologique sont hétérogènes, dans la mesure où ces instances appartiennent à différentes espèces biologiques.

Cette argumentation semble avoir des conséquences dramatiques : elle semble remettre en cause la réalité de *toutes* les espèces naturelles correspondant aux concepts des sciences autres que la physique fondamentale. Toute espèce qui n'est pas une espèce de la physique fondamentale regroupe des instances hétérogènes : elles sont hétérogènes du point de vue de toutes les sciences de niveau inférieur. Deux animaux d'espèces différentes qui perçoivent le même prédateur de type X (supposons qu'il s'agisse d'un prédateur pour les deux espèces) peuvent instancier la même espèce psychologique « perception d'un prédateur de type X » tout en étant dans différents états neurophysiologiques. Mais deux animaux (de même espèce) peuvent aussi instancier un même type d'état neurophysiologique, par exemple en ayant la même distribution d'activation neuronale dans une aire cérébrale donnée, tout en étant dans un état différent sur le plan des neurones individuels et surtout sur le plan chimique : la même aire cérébrale peut être active sans qu'il y ait identité sur le plan de l'activité des neurones individuels, et le même neurone peut être excité au même niveau sans que les mêmes canaux ioniques soient dans le même état dans les deux animaux. De même, deux substances peuvent appartenir à la même espèce chimique sans être de même nature physique : un échantillon d'eau peut contenir plus d'isotopes lourds de l'hydrogène que l'autre, ce qui est une différence physique sans être une différence chimique. Deux échantillons d'un gaz peuvent occuper le même état sur le plan des grandeurs thermodynamiques (de pression, température, entropie etc.), tout en étant dans différents états sur le plan microphysique : ils peuvent différer à l'égard de l'état de mouvement des molécules individuelles. L'omniprésence de la multiréalisabilité se retourne contre la stratégie argumentative de Kim : si la multiréalisabilité suffisait pour remettre en cause le statut scientifique des espèces multiréalisées, les seules espèces dont la réalité ne ferait pas de doute seraient les espèces de particules élémentaires, ainsi que leurs propriétés : masse, charge électrique, spin, etc.

Kim est conscient de la possibilité de « généraliser » la thèse de la multiréalisabilité au-delà de la physiologie des espèces biologiques. « Peut-être, dit-il, les espèces biologiques,

²³ Kim appelle cela le « principe de l'héritage causal » (p. 355).

comme on les conçoit normalement, sont-elles trop larges pour donner lieu à des bases de réalisation physiques-biologiques uniques ; la base neuronale de la douleur pourrait peut-être changer même dans un organisme individuel au cours du temps » (p. 313). Le problème est que la « généralisation » parallèle de son argument contre la réalité des espèces multiréalisables remet en cause l'existence des « psychologies spécifiques à l'espèce » (p. 334) tout autant que l'existence de la psychologie générale, c'est-à-dire de la psychologie qui vaut indépendamment de l'espèce biologique. Les espèces réelles, dit-il, sont des « types de structure physique » (p. 313). Mais la logique de son argument contre la réalité des espèces multiréalisables conduit à la thèse plus forte selon laquelle les seules espèces réelles sont des « types de structure physique *homogène* » qui ne sont pas eux-mêmes multiréalisables, autrement dit des espèces identifiés par la physique fondamentale des particules élémentaires.

Kim cherche pourtant à contenir la portée de sa conclusion, pour la restreindre à la seule irréalité des espèces de la psychologie générale (non spécifique aux espèces). « La doctrine [défendue ici] permet, et en fait encourage les 'psychologies spécifiques aux espèces' » (p. 334). Celle-ci, « par exemple la psychologie humaine [...] peuvent toutes s'épanouir en tant que théories scientifiques » (p. 335). Une manière de justifier ce traitement inégal de la multiréalisabilité aux différents niveaux est de soutenir que l'hétérogénéité de la base de réalisation est plus forte dans le cas de la psychologie générale que dans celui des espèces identifiées par des sciences de niveaux inférieurs à la psychologie. Une autre stratégie, que Kim adoptera plus tard²⁴, est de faire valoir qu'il existe une différence de nature logique entre la multiréalisabilité des états psychologiques par des états neurophysiologiques et la multiréalisabilité que l'on trouve à d'autres niveaux. Dans le cas de la psychologie, il s'agit d'une différence d'ordre *logique* : un état psychologique correspond à un prédicat de second ordre, alors que les prédicats qui font référence aux états neurophysiologiques qui le réalisent, sont de premier ordre. En revanche, aux autres niveaux, il s'agit de propriétés de premier ordre. La première stratégie ne semble pas prometteuse : nous avons déjà fait remarquer que l'hétérogénéité des réalisateurs neurophysiologiques des états cognitifs apparaît, à la lumière de la recherche récente en neurosciences cognitives, assez modeste. En ce qui concerne la seconde stratégie, elle semble également difficile à poursuivre : la seule raison d'accorder à un prédicat donné le statut de prédicat de premier ordre, et donc de prédicat faisant référence à une propriété réelle capable de faire une différence causale,

²⁴ J. Kim, « Does the Problem of Mental Causation Generalize ? », *Proceedings of the Aristotelian Society* 97 (1997), pp. 281-297, et *Mind in a Physical World*.

semble être le fait que cette propriété apparaisse dans des lois de la nature²⁵. Or, comme nous l'avons déjà fait remarquer, certaines branches de la psychologie générale (non spécifique à une espèce biologique), notamment la psychologie de la perception ou de l'apprentissage par conditionnement, sont parvenues à découvrir des lois ; nous avons donc des raisons de considérer que les propriétés qui figurent dans ces lois sont réelles et causalement efficaces. C'est en tout cas le critère qu'il faut retenir dans le cadre de l'analyse traditionnelle de la causalité en termes d'explication DN, et Kim n'a fait aucune tentative de le remplacer comme critère de l'identification des pouvoirs causaux réels.

²⁵ Dans « Multiple Realization, Reduction and Mental Properties », *International Studies in the Philosophy of Science* 13 (1999), p. 135-149, j'ai défendu la thèse selon laquelle c'est la présence dans des lois de la nature qui constitue le critère le plus adéquat pour juger si une propriété est naturelle et causalement efficace. Je montre premièrement que ce jugement ne dépend pas de l'homogénéité ou hétérogénéité de sa base de réduction et deuxièmement que la confirmabilité d'une hypothèse ne peut pas servir de critère du caractère naturel et efficace des propriétés qu'elle mentionne, dans la mesure où la confirmabilité dépend elle-même de la présence des propriétés qu'elle mentionne dans d'autres lois. Je montre que, à défaut d'accepter ce critère, la multiréalisabilité conduit à nier le caractère réel et causalement efficace de toutes les propriétés en dehors des propriétés reconnues par la physique fondamentale. G. Witmer, « Multiple Realizability and Psychological Laws », in : S. Walter et H.-D. Heckmann, *Physicalism and Mental Causation*, Exeter (U.K.), Imprint Academic, 2003, p. 59-84, a aussi argumenté que la multiréalisabilité d'une propriété n'empêche ni la confirmabilité d'une hypothèse qui la fait intervenir ni sa capacité de figurer dans des lois de la nature.