

Jaegwon Kim

L'esprit dans un monde physique

Essai sur le problème corps-esprit et la causalité mentale

Préface de Max Kistler

Traduit de l'américain par

François Athané et Édouard Guinet

Mind in a Physical World, An Essay on the Mind-Body Problem and Mental Causation

©1998, MIT Press, Cambridge (Mass.), ISBN

Première édition : Éditions Syllepse, septembre 2006

Le prestige et l'influence de Jaegwon Kim sur la philosophie contemporaine de langue anglaise ne peuvent guère être surestimés. Kim est un penseur de tout premier plan en métaphysique¹ et en philosophie de l'esprit, tout particulièrement en ce qui concerne le problème du rapport entre le corps et l'esprit. Nombre de concepts et d'arguments qu'il a mis en avant ont contribué à forger le cadre dans lequel la communauté philosophique de tradition analytique pose ce problème. La force et la clarté de sa réflexion sont tels que, pour s'y attaquer, depuis au moins une trentaine d'années, il est incontournable de tenir compte des thèses et arguments qu'il avance. L'une des raisons qui expliquent sa réputation d'extrême rigueur et de probité intellectuelle est son humilité devant la force des arguments : il a montré qu'il est prêt à remettre en question une thèse qu'il a lui-même longuement développée et défendue, dès lors qu'il trouve de puissants arguments à son encontre. Cela s'est produit notamment à l'égard du concept de survenance (*supervenience*) : convaincu de son importance pour l'analyse du rapport entre corps et esprit, Kim a consacré de nombreux travaux à l'élaboration de ce concept et à l'exploration de ses différentes variantes. Néanmoins, lorsqu'il a découvert des arguments qui montraient que la thèse de la survenance (des propriétés mentales sur les propriétés physiques – j'y reviendrai) est si faible qu'elle est compatible avec de nombreuses doctrines métaphysiques, dualistes aussi bien que matérialistes, il est revenu sur son jugement. Estimant

¹. Dans la tradition de philosophie analytique à laquelle appartient Kim, la légitimité de la métaphysique a fait l'objet d'un débat controversé. J'y reviendrai plus loin. Dans les années 1930, l'empirisme logique – l'ancêtre de la philosophie analytique d'aujourd'hui – soutenait qu'il n'y a pas de métaphysique rationnelle, la tâche de la philosophie étant uniquement de clarifier les concepts et formes de raisonnement à l'œuvre dans les sciences. Aujourd'hui, il y a un large consensus pour accepter le fameux argument de Quine [1951] selon lequel la distinction entre jugements analytiques – dont la vérité est exclusivement fondée sur leur forme – et synthétiques – dont la vérité est fondée sur l'expérience – est une question de degré. On en conclut qu'il est impossible de séparer la réflexion *a priori* sur les concepts et formes de raisonnement, d'une part, de la recherche scientifique et empirique sur les faits, d'autre part. Cela conduit à redonner un rôle légitime et nécessaire à la métaphysique : celle-ci a pour vocation d'accueillir toute réflexion sur le réel menée dans des catégories plus générales que celles utilisées par les différentes sciences. À titre d'exemples, les réflexions sur le rapport entre une substance et ses propriétés, sur la persistance des substances et des propriétés dans le temps, sur l'espace et le temps eux-mêmes, sur la causalité, sur les rapports de détermination entre les parties microscopiques d'un objet et ses propriétés macroscopiques, sont de nature plus conceptuelle et plus générale que des recherches scientifiques sur tel ou tel domaine de la réalité. Cependant, ces catégories sont indispensables, dans la mesure où elles sont constitutives du cadre conceptuel à l'intérieur duquel nous construisons une conception scientifique du monde, dont le contenu provient des résultats des différentes sciences. Étant donné cet objectif, il est clair que la réflexion métaphysique est conditionnée par la compatibilité de ses résultats avec les connaissances scientifiques. Une telle conception d'une métaphysique respectueuse des sciences est bien exposée in LOWE [1998].

désormais que la relation de survenance est métaphysiquement « superficielle », et donc inadaptée à nous éclairer sur le fond du rapport entre corps et esprit, Kim a emprunté de nouvelles directions de recherche. C'est l'un des épisodes qui ont fait apparaître l'énorme influence de la pensée de Kim : les raisons qui ont convaincu Kim ont également donné une nouvelle direction à l'ensemble du débat. J'y reviens un peu plus loin.

LA PREMIERE PHILOSOPHIE ANALYTIQUE ET LE « BEHAVIORISME LOGIQUE »

Sa trajectoire personnelle, au départ fort originale, est désormais représentative de l'évolution générale de la philosophie de tradition analytique, dont Kim est devenu l'un des acteurs majeurs. Il a notamment été à l'avant-garde d'un regain d'intérêt notable de la philosophie analytique pour la métaphysique. Né en Corée en 1934, Kim y entame des études de littérature française. Peu après la guerre de Corée, en 1955, il a l'intention de poursuivre ses études en France, mais les circonstances veulent qu'il finisse par rejoindre les États-Unis. Il a alors 20 ans et parle peu l'anglais. Pour sa thèse de doctorat en philosophie, il choisit Carl G. Hempel qui est, avec Carnap, Reichenbach et Feigl, l'un des principaux responsables de l'implantation durable du style dit « analytique » en philosophie dans les pays de langue anglaise, après la Seconde Guerre mondiale. Ce courant a son origine dans le Cercle de Vienne autour de Schlick, Carnap et Neurath, ainsi que le groupe de « philosophie scientifique » autour de Reichenbach à Berlin. Ils se proposaient de donner à la réflexion et au débat philosophiques une rigueur qu'on croyait réservée à la science, et ce, grâce à l'usage systématique de la nouvelle logique formelle, développée par Frege, Russell et Whitehead. Les questions métaphysiques traditionnelles, dont celle du rapport entre corps et esprit, étaient alors considérées comme de faux problèmes ou, comme le dit le titre de l'un des ouvrages de Carnap, des « pseudo-problèmes ».

Pour les philosophes du cercle de Vienne, ces faux problèmes ont leur source dans la trop grande importance métaphysique que l'on accorde aux structures grammaticales et lexicales des langues naturelles : étant donné que la phrase « cinq est un nombre premier » a la même structure grammaticale que la phrase « Paris est une ville » ou « Frege est un logicien », il semble naturel de supposer que le nombre cinq est un objet individuel, autant que la ville de Paris et Gottlob Frege. Or la ressemblance syntaxique des mots qui désignent respectivement Frege, la ville de Paris et le nombre cinq semble être une raison bien superficielle de conclure que le nombre cinq appartient à la catégorie ontologique des « objets individuels », comme Frege et la ville de Paris.

De manière semblable, la structure grammaticale de la phrase « Sa crainte d'une tempête, son désir d'éviter le danger et sa croyance qu'une tempête était imminente ont conduit à la décision de Pierre de rentrer au port » suggère qu'il existe des états et des événements dans l'esprit de Pierre qui s'enchaînent de manière tout à fait analogue aux états et événements physiques décrits dans la phrase : « Étant donné la déforestation de la pente et l'abondance des pluies, le débordement de la rivière a provoqué un glissement de terrain. » À ses débuts, l'un des projets principaux du nouveau courant « analytique » – qui ne portait pas encore ce nom, mais plutôt celui d'« empirisme logique » ou de « philosophie scientifique » – était de faire apparaître comme telle l'erreur consistant à se laisser abuser par la similitude des formes grammaticales. Il s'agissait de poser le problème de la nature métaphysique, par exemple, des nombres et des

états et événements mentaux, en comparaison avec les états et événements physiques. Or ce n'est pas parce qu'il y a le nom « cinq » et l'expression nominale « la décision de Pierre de rentrer au port » qu'il existe nécessairement un objet ou événement qui leur correspondent d'une manière analogue à l'objet correspondant au nom « Paris » et à l'événement correspondant à l'expression nominale « le débordement de la rivière ». Plutôt que d'accepter le présupposé de l'existence des nombres, des états et des événements mentaux, comme nous le suggère la grammaire de la langue naturelle, et de poser ensuite la question typiquement métaphysique de la nature ou de l'essence de ces entités, il s'agissait pour les empiristes logiques de « dépasser la métaphysique par l'analyse logique du langage », comme le dit le titre d'un fameux article de Carnap. À l'égard de l'esprit, l'enjeu était de montrer que les énoncés faisant apparemment référence aux entités mentales pouvaient être traduits en d'autres énoncés qui ne faisaient référence qu'à des entités observables. Une fois cette équivalence montrée, l'apparence de mystère quant à la « nature » de ces entités et de l'esprit dans son ensemble serait dissipée. Le fait que l'on découvre au contraire qu'une phrase n'a aucun équivalent ne faisant référence qu'à des entités observables, signifie qu'elle est démasquée comme dépourvue de signification cognitive : elle n'a pas de valeur de vérité objective. Cela n'empêchait pas Carnap de reconnaître aux phrases portant, par exemple, sur l'Absolu une valeur poétique comparable à celle des phrases de la poésie dadaïste : sans signification au sens où il serait inapproprié de demander si elles sont vraies ou fausses, elles peuvent néanmoins procurer un plaisir esthétique ou susciter des émotions². À l'égard du discours sur l'esprit, la conception vérificationniste que je viens d'esquisser mène au béhaviorisme logique : dans la mesure où les termes du langage mental sont significatifs, ils sont équivalents à des expressions logiquement construites à partir d'expressions observationnelles et de connecteurs logiques. Ainsi, pour donner un exemple utilisé par Hempel [1935], l'un des protagonistes les plus illustres de l'empirisme logique venu d'Autriche et d'Allemagne aux États-Unis pendant l'époque nazie³, la signification de l'énoncé « Paul a mal aux dents » consiste dans l'ensemble des conditions de sa vérification, exprimées dans des énoncés observationnels, tels que « Paul pleure et fait des gestes de telle ou telle espèce », « À la question : “Qu'est-ce que tu as ?” Paul articule les mots : “J'ai mal aux dents”, ... », *in* Fisette & Poirier [2002, p. 204].

LE RENOUVEAU DE LA CONCEPTION METAPHYSIQUE DU PROBLEME CORPS-ESPRIT

Cependant, le béhaviorisme logique a été abandonné, comme d'ailleurs plus généralement la forme radicale du vérificationnisme qui ne reconnaît comme termes significatifs que les expressions qui sont traduisibles en un pur langage observationnel. Il s'avère que, en psychologie comme dans les sciences physiques, il est possible de construire des explications plus simples et plus puissantes si on a recours aux termes théoriques, et ce seulement si l'on utilise des énoncés dont la signification n'est pas épuisée par une liste finie de leurs conséquences observables. La liste des conséquences observables possibles du fait que Paul a

². Selon CARNAP [1931], de nombreux énoncés de la métaphysique traditionnelle appartiennent à cette catégorie d'énoncés dénués de signification cognitive. Plus tard, CARNAP [1956] a montré que l'ontologie pouvait faire l'objet d'une analyse en termes de choix conventionnel d'un langage.

³. L'émigration forcée des philosophes de l'empirisme logique de l'Allemagne nazie, et plus tard de l'Autriche annexée par l'Allemagne, a conduit à implanter ce courant de manière durable dans les pays de langue anglaise et notamment aux États-Unis.

mal aux dents, comme de n'importe quel autre événement ou état mental, est ouverte et en principe infinie : le mal de dents peut pousser Paul à hurler mais aussi à réciter des poèmes ou à aller à la piscine. L'ensemble de ses autres états et événements mentaux contribue à déterminer les effets de son mal de dents sur son comportement. Comme l'a montré Hempel (1958 et Hempel 1965b) lui-même, c'est précisément leur relative indépendance par rapport à leurs conséquences observables qui fait l'intérêt et la fécondité de l'introduction de termes qui ne sont pas directement observationnels – autrement dit, de « termes théoriques ». Le maître de Kim faisait donc ici preuve de la même probité intellectuelle que lui : après avoir défendu le béhaviorisme logique, il a été, dès qu'il s'était convaincu des arguments puissants à son encontre, parmi ses plus fervents détracteurs.

Sous la direction de Carl Hempel, Jaegwon Kim choisit l'explication scientifique comme sujet de sa thèse de doctorat, un sujet parfaitement en phase avec l'attitude antimétaphysique de la philosophie analytique de l'époque. Mais peu de temps après, à partir du milieu des années 1960, Kim commence à s'intéresser au problème du rapport entre corps et esprit. Ce thème métaphysique des plus traditionnels avait été abordé par les empiristes logiques, et notamment par Hempel lui-même, dans la manière canonique de la première tradition analytique : par le biais de l'analyse linguistique du langage mental et de la signification des termes mentaux, le but étant de « dissoudre » le problème en tant que problème métaphysique. Corps et esprit ne seraient que deux registres linguistiques que la philosophie aurait pour vocation d'analyser. Mais peu à peu les concepts et thèses avancés pour expliquer le rapport entre ces deux registres, mental et physique, finirent par faire réapparaître le problème du rapport entre corps et esprit comme un véritable problème métaphysique. Le livre de Kim, *L'esprit dans un monde physique*, constitue l'un des points culminants de cette évolution. Il montre d'une manière particulièrement poignante comment les concepts utilisés pour analyser le rapport entre les registres de langage mental et physique, notamment ceux de survénance, de réalisation et de réduction, sont désormais à nouveau employés pour analyser le rapport métaphysique entre le corps et l'esprit, ou, dans les termes utilisés par Kim, entre les *propriétés* mentales et physiques.

De cette manière, le débat contemporain reprend le débat sur le rapport entre corps et esprit, tel qu'il a été conduit dans la philosophie classique. La motivation principale du matérialisme contemporain est de surmonter la difficulté que l'explication de l'interaction entre substance pensante et substance étendue pose à Descartes. Le dualisme, qui postule l'existence de deux substances absolument hétérogènes, est très bien placé pour expliquer les différences apparentes entre le corps et l'esprit ; mais la thèse de l'hétérogénéité radicale des deux substances rend l'interaction inconcevable. Hobbes faisait valoir dans ses objections aux *Méditations* de Descartes, que la seule manière d'expliquer l'esprit est de le concevoir comme matériel⁴. La métaphysique matérialiste contemporaine reprend cette thèse à son compte, et se propose de l'élaborer avec des outils conceptuels et logiques nouveaux.

LA SURVENANCE

⁴ « Troisièmes objections de Hobbes contre les *Méditations métaphysiques* de Descartes, objection seconde sur la seconde méditation », in DESCARTES [1642, p. 600-602].

Le cheminement de la réflexion que Kim a menée sur le concept de *survenance* est exemplaire à cet égard. Ce concept a été introduit dans la discussion sur le rapport entre le corps et l'esprit par Donald Davidson, dans le cadre d'une approche qui porte clairement les marques de la « vieille » tradition analytique visant à dissoudre, ou du moins à atténuer, les problèmes métaphysiques en les transformant en problèmes portant sur des types de discours : selon le monisme anomal de Davidson, l'esprit est réel et interagit causalement avec le monde physique, dans la mesure où chaque événement mental qui est cause d'un événement physique – comme lorsque la décision de Pierre le conduit à changer de cap – ou effet d'un événement physique – comme lorsque la perception d'un nuage noir provoque en Pierre la croyance qu'une tempête est imminente – est identique à un événement physique. La thèse de l'identité des événements mentaux avec des événements physiques dissout le mystère de l'intégration des événements physiques dans le réseau des interactions causales ; l'esprit y trouve sa place, par l'intermédiaire de l'identité des événements mentaux avec des événements physiques. Mais que devient alors le vieux problème corps-esprit ? N'affirmer que cette identité reviendrait davantage à nier le problème plutôt qu'à avancer vers sa solution, dans la mesure où celui-ci a son origine dans notre conviction profonde, renforcée par toute la tradition philosophique, qu'il existe une différence radicale entre l'esprit et le corps. Sans aucun doute, cette différence radicale entraîne une différence tout aussi radicale entre les événements mentaux et physiques. Comment l'identité proposée par Davidson est-elle intelligible étant donné l'hétérogénéité radicale de ses termes ? Davidson résout – il vaudrait mieux dire dissout – ce problème en mettant l'hétérogénéité sur le compte d'une divergence entre deux types de discours : le discours mental est soumis à une contrainte de rationalité qui n'a aucun équivalent dans le discours physique. Il s'agit d'une différence dans les normes de ce qui compte comme une explication satisfaisante : pour expliquer un événement physique, c'est-à-dire, pour Davidson, un événement en tant qu'il est décrit par le vocabulaire physique, il faut montrer qu'il existe des lois de la nature et des conditions initiales (la description d'une situation concrète) dont l'événement en question est une conséquence. Cela signifie que l'on peut construire un argument – un ensemble d'énoncés – valide dont les prémisses consistent en une description des conditions initiales et en l'énoncé des lois pertinentes, et qui a comme conclusion la description (physique) de l'événement. L'explication d'un événement mental consiste elle aussi en un argument, et se situe donc toujours sur le plan d'un rapport entre énoncés ; la différence résidant dans le critère d'acceptabilité : l'explication physique est satisfaisante si elle montre que la conclusion est une conséquence nécessaire des conditions initiales et des lois de la nature ; alors que l'explication psychologique est satisfaisante lorsque les prémisses font apparaître comme *rationnelle* l'action, la croyance ou le désir décrit dans la conclusion. Le concept de *survenance* entre dans cette conception afin de rendre compte de l'intuition d'une asymétrie systématique et fondamentale entre les deux types de description : la description physique d'un événement donné (étant présupposé qu'il s'agit d'un événement mental qui a donc aussi une description en vocabulaire psychologique) détermine sa description mentale, alors que la réciproque n'est pas vraie. Voici une définition de la *survenance* portant sur des ensembles de prédicats :

L'ensemble de prédicats mentaux M survient (fortement) sur l'ensemble de prédicats P si et seulement si nécessairement pour tout prédicat M_i , si un objet x satisfait M_i au temps t , alors il existe un prédicat P_i , tel que x satisfait P_i à t et nécessairement, tout ce qui satisfait P_i à un moment donné satisfait M_i à ce moment-là.

Si par exemple Pierre a mal aux dents à t , alors la *survenance* des prédicats psychologiques sur les prédicats neurophysiologiques garantit qu'il existe un prédicat neurophysiologique vrai de

Pierre à l'instant t , qui est tel que nécessairement, tout objet qui satisfait ce prédicat neurophysiologique à un moment donné, a mal aux dents à ce moment-là. Cette manière de s'exprimer nous conduit vers une seconde manière de caractériser la survenance : deux objets – ou plutôt, dans notre cas, deux personnes – qui satisfont les mêmes prédicats neurophysiologiques à un moment donné, satisfont nécessairement également les mêmes prédicats psychologiques ; ou encore, plus brièvement : si les prédicats M surviennent sur les prédicats P, alors il ne peut y avoir aucune différence à l'égard des M sans une différence concomitante à l'égard des P.

Le qualificatif « fortement », dans l'analyse de la survenance donnée plus haut, fait référence à la seconde occurrence de « nécessairement » : la survenance faible n'est qu'une survenance de fait dont il n'est pas requis qu'elle vaille nécessairement, et en particulier qu'elle est une conséquence des lois de la nature. Voici une manière de faire apparaître la différence entre survenance faible et forte un peu plus clairement : si les M surviennent sur les P, alors il n'y a jamais de différence entre deux objets à l'égard de leurs prédicats survenants M, sans qu'il y ait de différence à l'égard des prédicats sous-jacents P. Admettons par exemple que les prédicats d'évaluation morale des actions surviennent sur les prédicats physiques. Il n'y a pas, dans ce cas, deux actions qui diffèrent quant à leur valeur morale, sans qu'il y ait de différence physique sous-jacente : l'une ne peut pas être bonne et l'autre mauvaise s'il s'agit de gestes parfaitement semblables dans des circonstances parfaitement semblables. Nous pouvons interpréter cette survenance de manière faible ou forte. Selon l'interprétation faible, le fait que les différences morales soient toujours accompagnées de différences physiques est un fait contingent concernant notre monde réel. Autrement dit, la survenance faible n'interdit pas la *possibilité* d'une différence morale sans différence physique. Cela est difficile à comprendre dans la mesure où nous tendons à interpréter la survenance de manière forte : il paraît plausible qu'il *ne peut pas* y avoir de différence morale sans différence physique. Voici un exemple artificiel mais simple de survenance faible sans survenance forte : admettons qu'il n'y ait en France, à un moment donné, ni deux personnes qui portent le même prénom ni deux personnes qui ont exactement le même poids. Les prénoms surviennent alors, dans cette population, sur les poids, dans la mesure où il n'y a jamais de différence à l'égard du prénom sans différence de poids. Cependant, cette survenance n'est pas forte : il s'agit d'une pure coïncidence ; et rien n'empêche la possibilité que deux personnes de prénom différent acquièrent, peut-être parce que l'un suit un régime, exactement le même poids. En d'autres termes, dans la mesure où la survenance n'est que faible et non forte, il est vrai mais non nécessairement vrai que toute différence à l'égard des prédicats survenants s'accompagne d'une différence dans les prédicats sous-jacents.

La survenance est censée rendre compte de la détermination de la description mentale par la description physique, dans la mesure où, pour toute différence psychologique entre deux individus, il existe nécessairement une différence physique sous-jacente ; l'intuition naturelle étant que le physique détermine le mental ou que le mental dépend du physique. Selon la conviction physicaliste largement partagée dans ce débat, cette dépendance implique notamment que l'on ne peut induire aucun changement psychologique dans une personne sans induire un changement physique sous-jacent dans son cerveau.

PREDICATS ET PROPRIETES

Lorsqu'on caractérise la survenance, on doit choisir entre deux manières de s'exprimer : ou bien l'on dit – comme nous l'avons fait jusqu'ici – que ce sont des *prédicats* qui surviennent sur d'autres prédicats, ou bien l'on dit que la survenance concerne les *propriétés* désignées par ces prédicats. Le terme « propriété » prend dans la métaphysique contemporaine la relève du terme « attribut » ; il s'en distingue par l'abolition de la distinction traditionnelle entre « mode » et « attribut », requise dans le cadre des doctrines dualistes. Au lieu de parler, à la manière des nominalistes, de prédicats survenants sur des prédicats, on peut parler, dans le cadre de la métaphysique réaliste, de propriétés survenant sur des propriétés. Pour les nominalistes, les prédicats ne désignent pas une réalité indépendante, alors que les réalistes supposent qu'un prédicat tel que « est humain » désigne la propriété réelle d'être humain.

Cependant, les choses sont un peu plus complexes que je ne les ai présentées jusqu'ici, dans la mesure où un nominaliste peut parfaitement accepter de parler de « propriétés » et non seulement de « prédicats ». Il l'acceptera dans la mesure où il donne au mot « propriété » un sens faible, pour ainsi dire purement verbal et sans implication ontologique, selon lequel tout prédicat désigne une propriété. Non seulement, le prédicat « est un proton » désigne la propriété d'être un proton, mais le prédicat « est un proton ou est blond et joue de la guitare » désigne lui aussi une propriété, à savoir la propriété « disjonctive » d'être un proton ou d'être blond et de jouer de la guitare. Avec cette acception faible du mot « propriété », lorsqu'on remplace, dans les caractérisations de la survenance en termes de descriptions et de prédicats, l'expression « satisfait un prédicat » par « possède une propriété », on n'introduit qu'une modification verbale sans changer le contenu. Cependant, il est désormais courant d'interpréter le mot « propriété » d'une manière plus forte : selon celle-ci, le prédicat « est un proton ou est blond et joue de la guitare » ne désigne pas de propriété « réelle » ou « naturelle », c'est-à-dire une propriété qui existe indépendamment des prédicats qui la désignent. Selon les métaphysiciens réalistes, il y a une différence objective entre les prédicats qui désignent des propriétés « naturelles », comme celle d'être un proton, et des propriétés purement nominales, parmi lesquelles figurent les propriétés disjonctives et négatives, par exemple la propriété de ne pas être un proton. En particulier, ces dernières ne sont pas nomologiques : aucune loi de la nature ne fait intervenir la propriété de ne pas être un proton⁵. Nous pouvons utiliser la distinction entre le sens faible et le sens fort du mot « propriété » pour interpréter le passage – au chapitre 4 du présent ouvrage – où Kim donne ses raisons de contester l'existence des « propriétés disjonctives » : elles n'existent pas au sens d'être nomiques, c'est-à-dire au sens d'être des termes de lois de la nature.

En revanche, Kim, comme déjà Davidson, caractérise la survenance en termes de propriétés possédées par les objets, et non en termes de prédicats que satisfont ces objets : il ne peut pas y avoir deux objets qui possèdent, à un moment donné, les mêmes propriétés physiques tout en étant différents à l'égard de leurs propriétés psychologiques. Dans ce contexte, la référence aux propriétés doit être interprétée de manière compatible avec le nominalisme, c'est-à-dire qu'elle ne comporte pas d'engagement pour l'existence « réelle », ou nomique, de ces propriétés.

⁵. Pour la distinction entre prédicats et propriétés, ou entre propriétés en général et propriétés naturelles, voir LEWIS [1983] ; MELLOR & OLIVER [1997].

LA FAIBLESSE METAPHYSIQUE DE LA SURVENANCE

Tout cela nous prépare à apprécier l'un des apports les plus significatifs de Kim à la réflexion sur le rapport corps-esprit. Après avoir travaillé pendant des années à la découverte de toute une série de concepts de survenance avec des différences subtiles, dans le but de trouver celui qui serait le plus adéquat à rendre compte du rapport entre propriétés physiques et psychologiques⁶, Kim s'aperçoit de la faiblesse explicative et métaphysique de ce concept : cette découverte est un indice clair de la transformation du problème corps-esprit en problème métaphysique. Cela signifie que l'on s'autorise à s'interroger sur le rapport entre le corps et l'esprit en tant que tels : la distinction entre le corps (et le cerveau) et l'esprit est présupposée par les registres conceptuels des sciences pertinentes, la neurophysiologie d'une part et la psychologie d'autre part. En même temps, elle n'est pas épuisée par l'ensemble des connaissances scientifiques : l'esprit, en tant qu'objet de l'investigation psychologique, n'est qu'incomplètement connu à un certain stade de l'histoire de cette science. La même chose vaut pour le corps et le cerveau, à l'égard de leur rapport à la science neurophysiologique. Cette dernière présuppose l'existence de son objet d'étude sans le connaître exhaustivement, à un moment donné de l'histoire. Les concepts métaphysiques d'esprit et de cerveau visent les objets de ces sciences, en tenant compte du caractère limité des connaissances effectives à leur sujet à un moment donné. Kim fait remarquer, et cela dès la postface du recueil réunissant ses travaux sur la survenance (Kim, 1993), qu'il s'agit d'un concept métaphysiquement superficiel : premièrement, et contrairement à la première intuition, la survenance n'implique en elle-même ni la dépendance des propriétés survenantes par rapport aux propriétés sous-jacentes, ni leur détermination par ces dernières. La survenance ne consiste en rien d'autre qu'une corrélation systématique entre deux ensembles de propriétés (ou prédicats), mais n'exerce absolument aucune contrainte sur l'origine de cette corrélation : les propriétés M peuvent survenir sur les propriétés P sans en être déterminées et sans en dépendre, notamment si les deux ensembles de propriétés sont déterminés par un troisième ensemble de propriétés qui agit à leur égard de manière analogue à une cause commune : supposons par exemple que l'ensemble D des propriétés – ou actes de volonté – divines détermine à la fois M et P, de telle sorte qu'il n'y a jamais deux objets qui diffèrent à l'égard des M sans également différer à l'égard des P. Dans ce cas, les M surviennent sur les P, sans en dépendre et sans être déterminées par elles. Dans une telle situation, toutes les instances de M et de P ne *dépendent* que des actes de volonté divine, D, et ne sont *déterminés* que par eux.

De manière semblable, le fait que le concept de survenance impose une contrainte asymétrique ne garantit pas pour autant que le rapport objectif entre les deux ensembles de propriétés soit lui-même asymétrique. Certes, il est vrai que le concept de survenance des M sur les P contient une asymétrie dans la caractérisation de la corrélation entre les deux ensembles de propriétés : elle implique qu'il ne peut pas y avoir de différence à l'égard des M sans différence à l'égard des P, alors qu'il peut très bien y avoir des différences à l'égard des P sans différence quelconque à l'égard des M. Cette dernière possibilité pourrait par exemple correspondre à un changement dans l'activité d'une seule cellule nerveuse cérébrale qui n'intervient dans la détermination des propriétés psychologiques que par l'intermédiaire de l'activité de nombreuses autres cellules, de

⁶ Le recueil *Supervenience and Mind* [Kim, 1993/2009, vol. 1 ; 2014, vol. 2] réunit quelques-uns des articles les plus importants que Kim a publié à ce sujet.

telle sorte que ce changement microscopique n'a pas la moindre conséquence au niveau psychologique. Mais on voit la faiblesse et le caractère non contraignant de cette asymétrie si l'on fait l'observation suivante : simplement, le fait que les M surviennent sur les P n'exclut pas que les P surviennent également sur les M, auquel cas on a affaire à une corrélation symétrique une à une. Cette possibilité confirme le fait, déjà constaté auparavant, que la survenance des M sur les P n'impose aucune contrainte sur les relations métaphysiques de détermination et de dépendance : il est clair que la situation où cette survenance est complétée par une survenance réciproque, est compatible avec l'existence de relations de dépendance et de détermination quelconques. La corrélation peut avoir son origine métaphysique dans le fait que les M déterminent les P, dans le fait que les P déterminent les M ou dans le fait qu'un troisième ensemble de propriétés D détermine à la fois les P et les M.

La découverte par Kim du caractère superficiel de la relation de survenance a considérablement infléchi le débat sur le rapport entre le corps et l'esprit. La survenance du mental sur le physique est toujours admise comme une contrainte minimale dont l'acceptation définit une position matérialiste largement partagée ; en ce sens, Kim la baptise « physicalisme minimal ». Cependant, on considère désormais qu'elle n'est pas apte à éclaircir l'énigme du rapport entre corps et esprit. Si nous supposons que le mental survient sur le physique, nous ne savons encore rien sur l'origine de cette survenance : par conséquent, la recherche philosophique s'est tournée vers d'autres concepts plus susceptibles de nous expliquer le pourquoi de cette survenance ; en d'autres termes, il s'agit d'élaborer des concepts plus aptes à apporter un éclairage métaphysique. Plusieurs de ces concepts occupent une place importante dans *L'esprit dans un monde physique*, notamment les concepts de réduction et d'émergence.

MULTIREALISABILITE ET REDUCTION FONCTIONNELLE

Là encore, le travail de Kim a considérablement infléchi le cours de la discussion : depuis une trentaine d'années, toutes les parties du débat prenaient pour acquis, premièrement, que les propriétés mentales sont, au moins en principe et en possibilité sinon en réalité, « multiréalisables », c'est-à-dire que deux systèmes cognitifs peuvent partager une propriété mentale sans partager de propriété neurophysiologique sous-jacente. Cette possibilité peut être admise dans le cadre de la conception « fonctionnaliste » des états et événements mentaux. Selon celle-ci, l'identité d'un événement mental est indépendante du substrat dans lequel il a lieu. Ce qui fait de ma croyance que le tapis dans cette pièce est rouge une instance de cette espèce d'état mental, et ce qui le distingue d'états semblables, comme le désir de remplacer le tapis dans cette pièce par un tapis bleu, ou la croyance que le tapis dans cette pièce est bleu, ce sont des relations causales : ces états sont causés, plus ou moins indirectement, par des perceptions ; ils causent, indirectement ou non, des actions, et ils entretiennent surtout de multiples et complexes interactions causales avec d'autres états et événements mentaux. Selon la conception fonctionnaliste, il est envisageable que des événements différents à l'égard de leurs propriétés physiques ou physiologiques partagent néanmoins leur « rôle » causal, c'est-à-dire l'ensemble de leurs interactions causales. Dans ce cas, ces événements sont différentes « réalisations » de la même espèce d'événement mental ; autrement dit, ce type d'événement mental est « multiréalisable ». La seconde thèse largement acceptée par les philosophes

fonctionnalistes est que cette multiréalisabilité est incompatible avec la *réduction* des propriétés psychologiques aux propriétés neurophysiologiques. Ce que Ned Block a appelé le « consensus antiréductionniste » était largement dû à la conviction que la réalisabilité multiple fournissait une raison concluante de penser que le mental était irréductible en principe au neurophysiologique.

Pour le même motif, ainsi que pour les raisons avancées par Davidson et mentionnées plus haut, on considérait comme définitivement établi que le concept adéquat pour rendre compte du rapport entre corps et esprit devait être *plus faible* que la réduction entre théories, telle que Nagel l'avait conçue⁷ : pour Nagel, une théorie ne peut être réduite à une autre que par l'intermédiaire de principes-pont qui expriment des relations de dépendance nomologique entre propriétés réduite et réductrice. Or la multiréalisabilité était considérée comme incompatible avec l'existence de tels principes-pont, dans la mesure où une propriété multiréalisable n'est pas corrélée avec une propriété unique, mais peut être corrélée avec de nombreuses propriétés diverses.

Kim a réussi à ébranler ces deux convictions : il a premièrement montré que la réduction à la Nagel n'est pas un concept trop restrictif pour être adéquat au rapport entre propriétés neurophysiologiques et psychologiques, mais au contraire un concept trop faible : de manière semblable à la survenance, il s'avère que la réduction par principes-pont est compatible avec plusieurs schèmes métaphysiques très différents et incompatibles entre eux. Pour être adéquat à une position physicaliste mais incompatible avec, par exemple, le dualisme des substances, le concept de réduction devant rendre compte du rapport entre propriétés physiologiques et psychologiques doit être plus fort que celui de Nagel : la réduction doit exprimer un lien plus étroit entre propriété réductrice et réduite que la simple existence de principes-ponts. Deuxièmement, Kim montre que son nouveau concept de « réduction fonctionnelle » satisfait à cette exigence, en imposant des contraintes plus fortes que la réduction de Nagel, tout en étant compatible avec la réalisabilité multiple.

Il fallait toute l'inventivité et la patience conceptuelle de Kim pour développer une telle conception, présentée dans *L'esprit dans un monde physique*. Il s'y emploie à dissoudre l'apparence paradoxale du projet de concevoir le rapport entre cerveau et esprit d'une manière à la fois plus forte qu'une conception déjà jugée trop forte (celle de Nagel) et plus faible qu'elle. La conception ici développée par Kim est plus forte qu'une réduction à la Nagel, dans la mesure où elle est, contrairement à cette dernière, incompatible avec les doctrines dualistes. En même temps, elle est plus faible qu'elle dans la mesure où elle est compatible avec la réalisabilité multiple que la réduction de Nagel excluait.

L'ESPRIT : CAPABLE D'INTERVENIR CAUSALEMENT DANS LE MONDE OU VOUE A L'ELIMINATION ?

Depuis sa première parution en 1998, *L'esprit dans un monde physique* a une nouvelle fois changé la donne du débat. Kim a poussé les avocats du « consensus antiréductionniste » dans leurs retranchements : son livre les met au défi de se défendre contre la conclusion de l'« argument de la survenance ». Contrairement à ce que nous croyons intuitivement, nos

7. NAGEL [1961], en particulier le chap. 11, intitulé « The Reduction of Theories ».

décisions, nos intentions, et tous les autres événements ayant lieu dans notre esprit ne sont pas eux-mêmes causalement responsables des mouvements corporels constitutifs de nos actions. Ces mouvements n'ont que des causes physiques. L'apparence de causes mentales de ces mouvements corporels vient du fait que nous utilisons des *concepts* psychologiques pour rendre compte de leur origine. Mais les événements qui tombent sous ces concepts et qui sont les causes de nos actions sont toujours des événements physiques, plus précisément neurophysiologiques. Ce résultat peut paraître décevant pour l'antiréductionniste qui espérait pouvoir concilier l'irréductibilité du mental avec son efficacité causale autonome et indépendante de l'efficacité du substrat physique. Mais le raisonnement de Kim montre qu'il a tort d'être déçu : seule leur réduction (fonctionnelle, selon Kim) peut garantir l'efficacité des propriétés et événements mentaux ; le prix de l'irréductibilité est l'inefficacité causale, et donc, en fin de compte, l'élimination. En revanche, la réduction fonctionnelle d'une propriété mentale justifie, premièrement, l'efficacité causale de chacune de ses instances – qui est, à chaque fois, également l'instance d'une propriété physique – et garantit, deuxièmement, l'autonomie du *concept* mental : le concept mental reste indispensable en tant qu'objet de la psychologie, même si le monopole de l'efficacité causale appartient aux propriétés physiques réalisatrices.

Kim reconnaît une seule exception à ce verdict : l'aspect qualitatif de l'expérience est bien réel, sans être pour autant susceptible d'une analyse fonctionnelle, et donc sans être réductible. Il y a un « effet que cela fait », pour utiliser l'expression forgée par Thomas Nagel (1979) qui est désormais couramment utilisée dans ce débat, de voir le rouge d'une rose, ou d'entendre une mélodie. Cet effet, souvent appelé « *quale* », est une propriété subjective de mon expérience qui est, selon Kim, irréductible à l'aspect fonctionnel des représentations sous-jacentes. Ce jugement sur les « *qualia* » – qui sont réels tout en n'étant pas susceptibles d'une réduction fonctionnelle – nuance le physicalisme soutenu par Kim, d'où le titre de l'ouvrage qu'il a publié à la suite de *L'esprit dans un monde physique : Physicalism or Something Near Enough*⁸.

La force des arguments de Kim, sa manière patiente, minutieuse et rigoureuse d'aller toujours jusqu'au bout des implications de ses raisonnements obligent tous les philosophes désireux de comprendre le rapport entre corps et esprit d'affronter la conclusion de Kim : à strictement parler, ce n'est pas notre esprit qui change le monde lorsque nous agissons, mais uniquement notre cerveau.

Max KISTLER

Université Paris 1 Panthéon-Sorbonne, IHPST (Institut d'histoire et de philosophie des sciences)

⁸. Kim [2005] : *Le Physicalisme, ou quelque chose qui y ressemble d'assez près*.