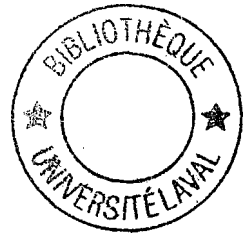
CHRISTIAN LACROIX

CAUSALITÉ MENTALE ET RÉDUCTIONNISME CHEZ JAEGWON KIM

Mémoire
présenté
à la Faculté des études supérieures
de l'Université Laval
pour l'obtention
du grade de maître ès arts (M.A.)

Faculté de philosophie
UNIVERSITÉ LAVAL

Novembre 1999



Causalité mentale et réductionnisme chez Jaegwon Kim

Résumé

Dans ses plus récents écrits, Jaegwon Kim soutient que seule une approche réductionniste est en mesure de rendre compte de la causalité mentale tout en respectant nos convictions physicalistes. Ce faisant, il va à l'encontre de la tendance actuelle en philosophie analytique de l'esprit représentée par le physicalisme non-réductif. Je tente ici d'évaluer si Kim réussit à faire du réductionnisme une approche valable et intéressante. Pour ce faire, je présente en détail et commente l'argument de Kim servant à réfuter le physicalisme non-réductif, de même que les deux modèles de réduction qu'il a élaborés. Je conclus que Kim réussit à réfuter le physicalisme non-réductif, mais que le réductionnisme qu'il propose ne présente qu'un intérêt limité puisqu'il laisse de côté les qualia. Aucune solution satisfaisante ne semble donc pouvoir être apportée au problème de la causalité mentale.

Christian Lacroix
Étudiant

Renée Bilodeau
Directrice

Avant-propos

La philosophie analytique n'est plus si jeune, mais même aujourd'hui son influence se fait à peine sentir dans certains endroits du globe... comme l'Université Laval. Cet état de fait est cependant en train de changer et je suis heureux d'y contribuer par l'ajout de ce mémoire à la collection de la bibliothèque.

La rédaction d'un premier texte philosophique de quelque envergure ne se fait pas sans peine, aussi j'aimerais dire merci à ceux qui l'ont partagée avec moi. J'aimerais d'abord remercier ma directrice, Renée Bilodeau, pour ses nombreux conseils judicieux et pour son temps si précieux, particulièrement cet été. J'aimerais également remercier mon épouse, Mélanie Frappier, avec qui j'ai vécu les hauts et les bas de la rédaction du mémoire de maîtrise, elle sur Heisenberg, moi sur Kim.

Christian Lacroix

Le 11 août 1999.

Table des matières

Causalité mentale et réductionnisme chez Jaegwon Kim

Introduction

Les états mentaux se divisent en deux groupes: les états mentaux phénoménaux et les états mentaux intentionnels. Les états mentaux phénoménaux, ou conscience phénoménale, désignent tous les états mentaux qui consistent en des effets qualitatifs et subjectifs produits sur notre esprit par nos perceptions, sensations et émotions. On désigne ces qualités subjectives par les expressions 'sense data', 'raw feel'(sensation brute) ou, plus couramment de nos jours, 'qualia'. L'impression de rouge que produit la vue d'une tomate mûre, l'effet que produit en nous l'odeur de la rose et la sensation de douleur que provoque une brûlure sont tous des exemples de qualia. Il n'existe pas de définition des qualia qui ne soit pas circulaire. Outre par le biais d'exemples, on réfère aux qualia en disant qu'ils correspondent à "ce que cela fait"que d'être dans tel état mental (d'après l'expression "*what is it like*" de Thomas Nagel, 1974). Les états mentaux intentionnels, quant à eux, désignent les croyances, les désirs, les intentions, etc., c'est-à-dire ce qu'on exprime par des énoncés propositionnels du genre "Je crois que...", "Il craint que...", "Jean pense que...", etc. Ce sont les états mentaux intentionnels qui entrent dans nos raisonnements et constituent la base de la rationalité. Les émotions constituent un cas mixte: elles possèdent à la fois un aspect phénoménal (ce que cela fait d'avoir peur ou d'être en amour, etc.) et un aspect intentionnel (peur de quelque chose, amour de quelque chose, etc.).

Nous référons couramment aux phénomènes mentaux pour expliquer nos comportements et ceux des autres. Par exemple, nous pouvons expliquer le fait que Julie immobilise sa voiture au feu rouge par le fait qu'elle connaît les règles du code de la route (croyances), qu'elle veut les respecter (intention ou désir) et qu'elle a vu que le feu était rouge (perception). Mais pour que les états mentaux puissent constituer la *cause* du comportement de Julie, il est nécessaire que les états mentaux *possèdent des pouvoirs causaux*. Comment concevoir que des états mentaux comme des qualia, des croyances, des désirs, etc. agissent sur le monde? Rien dans

la caractérisation que j'ai donnée des états mentaux ne laisse présager que ceux-ci sont à même d'interagir avec le monde physique. C'est là le *problème de la causalité mentale*.

Il faut noter cependant que la question en litige n'est pas de savoir *si* les états mentaux possèdent ou non des pouvoirs causaux. En effet, on *assume* que les propriétés mentales possèdent effectivement des pouvoirs causaux. Si ce n'était pas le cas, tant la psychologie scientifique que la 'psychologie' informelle que nous employons quotidiennement pour expliquer nos propres comportements et ceux des autres (*'folk psychology'*) seraient complètement erronées; une conséquence que très peu de philosophes seraient prêts à accepter. La question est par conséquent de savoir *comment* cela est possible, c'est-à-dire de donner une explication satisfaisante de la manière dont 'l'esprit' agit sur le corps.

Le problème vient de ce que nous ne pouvons concevoir l'interaction de choses n'ayant aucune propriété en commun. Déjà, Descartes avait bien des ennuis à faire accepter à ses contemporains que deux substances aussi différentes l'une de l'autre que l'âme, substance immatérielle, et le corps, substance matérielle, puissent interagir. Ainsi la princesse Élisabeth de Bohème écrivît à Descartes: "j'avoue qu'il me serait plus facile de concéder la matière et l'extension à l'âme, que la capacité de mouvoir et d'être ému, à un être immatériel."[1]

Certes, le dualisme des substances est largement abandonné aujourd'hui, mais ce genre de problème s'applique apparemment aussi bien au dualisme des *propriétés* qui possède de nombreux partisans de nos jours. En effet, comment concevoir l'action d'un événement *x*, *en tant qu'il possède la propriété mentale M*, sur un événement physique *y*? Est-ce que cela ne ressemblerait pas à de la magie ou à de la télékinésie, bref, à quelque chose de *non-naturel*? Une solution, bien sûr, est de suivre la voie que propose Élisabeth: rendre le mental physique d'une façon ou d'une autre. Le problème de la causalité est alors réglé, puisqu'il n'y a aucun problème à attribuer des pouvoirs causaux à un événement physique. Cette approche est d'ailleurs celle du *réductionnisme*. La forme traditionnelle du réductionnisme (J. J. C. Smart et H. Feigl) est de poser l'identité des différents types d'états mentaux avec des états physiques,

[1] Lettre de juin 1643, Descartes, Oeuvres Choisies, Tome II, p. 120.

généralement des états neurologiques du cerveau. La "vraie nature" de nos états mentaux se révèle donc être physique, ce qui permet au mental d'être investigué et expliqué selon les mêmes principes que ceux utilisés pour investiguer et expliquer le reste du monde physique. Par contre, s'agit-il vraiment de causalité mentale *qua* mentale? Et qu'advient-il de la conscience phénoménale dans une telle optique de réduction? Il est difficile d'imaginer pouvoir préserver l'aspect "qualitatif et subjectif" du mental en l'identifiant avec quelque chose d'observable à la troisième personne. Mais si l'on se refuse à opérer une telle réduction, est-ce que l'on peut vraiment montrer comment le mental exerce des pouvoirs causaux?

La plupart des philosophes travaillant sur le problème corps-esprit croient que oui, puisqu'ils s'inscrivent dans ce que l'on appelle le "consensus du physicalisme non-réductif". Le physicalisme non-réductif est une position qui tente justement d'accorder des pouvoirs causaux aux états mentaux tout en s'opposant à toute réduction des propriétés mentales aux propriétés physiques. On parle d'un consensus du physicalisme non-réductif parce qu'il ne s'agit pas d'une approche en soi, mais plutôt d'un regroupement de plusieurs approches, concurrentes par ailleurs, mais qui s'accordent sur quelques thèses considérées primordiales, comme le physicalisme et le réalisme du mental.

Bien que sympathique au physicalisme non-réductif, Jaegwon Kim a consacré plusieurs écrits destinés à montrer l'*incompatibilité* du physicalisme et du réalisme du mental. Pour ce faire, Kim présente deux principaux arguments, *le problème de l'exclusion causale/explicative* et *l'argument de la survenance*, qui portent à croire que les physicalistes n'ont d'autres choix que d'accepter le réductionnisme ou l'épiphénoménisme. Or, tant le réductionnisme que l'épiphénoménisme s'opposent aux thèses du physicalisme non-réductif. En fait, tant le réductionnisme que l'épiphénoménisme nous conduisent à la perte du mental tel que nous le concevons présentement. D'où un dilemme: comment choisir entre deux mauvaises options? Voici comment Kim exprime ce dilemme dans *Philosophy of Mind* (1996):

> Nous nous trouvons donc devant un difficile dilemme: Si nous sommes prêts à adopter le réductionnisme, nous pouvons expliquer la causalité mentale. Toutefois, en réduisant la capacité mentale aux propriétés physiques et/ou biologiques, nous pourrions bien perdre le caractère subjectif et intrinsèque de notre esprit - dont on peut soutenir qu'il

> constitue ce qui rend le mental, mental. En quel sens, alors, avons-nous sauvé la causalité "mentale"? Mais si nous rejetons le réductionnisme, nous ne voyons pas comment la causalité mentale peut être possible. Sauver le mental en perdant la causalité ne semble pas revenir à sauver quelque chose qui mérite d'être sauvé. Car que vaut l'esprit s'il n'a pas de pouvoirs causaux? D'une façon ou de l'autre, nous risquons de perdre le mental. Tel est le dilemme.[2]

L'existence de propriétés mentales qui soient irréductibles aux propriétés physiques mais dépourvues de pouvoirs causaux, c'est ce que propose l'*épiphénoménisme*. Autrement dit, pour Kim, le dilemme est que nous devons choisir entre le réductionnisme et l'épiphénoménisme, deux positions qui enlèvent quelque chose de fondamental à notre conception du mental. De fait, aucune de ces deux options n'a de quoi réjouir les partisans du physicalisme non-réductif. D'où le dilemme.

J'ai soulevé la question de savoir comment on pouvait choisir entre deux mauvaises options. La réponse est, bien sûr, en prenant la moins mauvaise. C'est ce que propose Kim qui, malgré sa sympathie pour le physicalisme non-réductif, porte aussi un intérêt marqué au réductionnisme, qui constitue son premier contact avec le problème corps-esprit. On peut déceler cet attrait dans ce passage de Kim extrait d'une rétrospective des courants philosophiques ayant dominé le débat corps-esprit:

> Pour plusieurs d'entre nous qui, comme moi, ont fait leurs études supérieures à la fin des années 50 et au début des années 60, le matérialisme de Smart et de Feigl représente notre premier contact avec le problème corps-esprit en tant que problème philosophique systématique. Leur approche apparaissait remarquablement courageuse et ingénieuse, de même qu'en accord avec l'optimisme scientifique de l'époque. C'était une idée intrigante et excitante que les événements mentaux pourraient simplement *être* des processus du cerveau et que la recherche scientifique pourrait le démontrer, tout comme

---

[2] *"So we find ourselves in a profound dilemma: If we are prepared to embrace reductionism, we can explain mental causation. However, in the process of reducing mentality to physical/biological properties, we may well lose the intrinsic, subjective character of our mentality - arguably, the very thing that makes the mental mental. In what sense, then, have we saved "mental" causation? But if we reject reductionism, we are not able to see how mental causation should be possible. But saving mentality while losing causality doesn't seem to amount to saving anything worth saving. For what good is the mind if it has no causal powers? Either way, we are in danger of losing mentality. That is the dilemma."* Kim (1996), p. 237. On retrouve également l'équivalent de ce dilemme sous d'autres formulations dans Kim (1993a) pp. 366-367, Kim (1998) pp. 118-119, de même que dans Macdonald et Macdonald (1995a) pp. 4-5.

la science nous a montré que la lumière est une radiation électromagnétique et les gènes des molécules d'A.D.N.[3]

De ce simple intérêt, Kim est passé à la défense du réductionnisme comme on le verra au cours de ce mémoire. Ce n'est pas dire que Kim propose de retourner à l'approche de Smart et Feigl. Le réductionnisme traditionnel fait face à des difficultés que Kim reconnaît. De plus, il partage la plupart des intuitions des partisans du physicalisme non-réductif. Seulement, il considère que le physicalisme non-réductif ne peut rendre compte de l'attribution de pouvoirs causaux aux états mentaux. Et à choisir entre l'épiphénoménisme et le réductionnisme, Kim choisit le réductionnisme sans hésitation. Un réductionnisme amélioré qui contournerait les principales difficultés auxquelles fait face le réductionnisme traditionnel, mais un réductionnisme tout de même. Kim écrit: "J'espère vous persuader que le réductionnisme dans le cas de l'esprit est une position philosophique sérieuse et motivée, et que même si nous pouvons en fin de compte décider de le rejeter, nous devrions le faire pour les bonnes raisons."[4]

La défense du réductionnisme n'est pas une tâche facile: si une approche au problème corps-esprit se place à contre-courant, c'est bien celle-ci. Le réductionnisme classique a connu une chute rapide qui a laissé un vif souvenir dans la mémoire des philosophes, les arguments qui ont provoqué cette chute apparaissent toujours aussi valides et le mot 'réductionnisme' a aujourd'hui une connotation naïve et simpliste qui entache inévitablement tout partisan de cette option.

Cependant, le succès ou l'échec de Kim à faire du réductionnisme un candidat sérieux repose avant tout sur la force des arguments qu'il avance. Dans ce mémoire, je vais donc examiner ces arguments et tenter d'en mesurer la force et la pertinence. Je soutiendrai que l'argument

---

[3] "*For many of us who, like me, went to graduate school in the late 1950s and early 1960s, Smart's and Feigl's materialism was our first encounter with the mind-body problem as a systematic philosophical problem. Their approach sounded refreshingly bold and tough-minded, and seemed in tune with the optimistic scientific temper of the times. It was an intriguing and exciting idea that mental events could just be brain processes, and that scientific research could show this, just as science showed us that light was electromagnetic radiation, and the gene were DNA molecules*" Kim (1998) p. 2.

[4] "*I hope to persuade you that reductionism about the mind is a serious, motivated philosophical position, and although in the end we may decide to reject it, we should do so for the right reasons.*" Kim (1998), p.89.

de la survenance de Kim représente une menace sérieuse pour le physicalisme non-réductif. Je montrerai cependant que son principe de l'exclusion explicative, sur lequel Kim base en grande partie l'argument de la survenance, est plus discutable. Heureusement pour Kim, le principe de l'exclusion causale, lui, apparaît indiscutable et c'est sur ce principe que repose en fait l'argument de la survenance. Je présenterai ensuite les deux modèles réductifs successifs que Kim a élaborés afin de fournir une alternative aux approches du consensus du physicalisme non-réductif. Je m'attarderai plus longuement sur le second modèle, celui de la réduction fonctionnelle, puisqu'il s'agit de celui que défend Kim actuellement. Je montrerai que ce modèle permet d'effectuer la réduction de certains états mentaux, mais que, comme toutes les approches réductionnistes, il ne permet pas de rendre compte de la causalité mentale. De plus, ce modèle n'est pas en mesure de réduire les qualia, ce que Kim reconnaît. Pour expliquer les qualia, Kim se tourne vers l'émergentisme, mais je soutiendrai que cette approche présente encore des problèmes et qu'une version moderne de l'émergence, celle de Humpreys, conduit possiblement soit à l'abandon du physicalisme, soit à l'abandon des propriétés mentales.

Plan détaillé du mémoire

Je commencerai par présenter, au premier chapitre, certaines notions fondamentales dont j'aurai besoin au cours de cet essai. J'exposerai d'abord ce que l'on entend par fermeture du monde physique et je présenterai le consensus du physicalisme non-réductif et ce qu'il implique. J'en profiterai également pour exposer brièvement les deux principales approches qui se retrouvent sous le large étendard du physicalisme non-réductif, soit le monisme anomal et le fonctionnalisme, de même que celles qui s'y opposent, soit le réductionnisme, l'éliminativisme et l'épiphénoménisme. Je parlerai ensuite des qualia et je présenterai trois 'arguments' mettant en évidence leurs principales caractéristiques et les raisons de les accepter à l'intérieur d'une théorie complète du mental. Finalement, je présenterai la survenance (de l'anglais *supervenience*) à laquelle Kim a consacré plusieurs articles. Il s'agit d'une notion importante puisque, dans les vingt dernières années, elle a souvent été perçue comme la solution idéale aux problèmes de la causalité mentale et des qualia. Elle est également essentielle à l'argument de la survenance que je présenterai au chapitre 2. Après avoir exposé

les principaux types de survenance, soit la survenance faible, la survenance forte et la survenance globale, je montrerai également qu'à y regarder de plus près la relation de survenance n'est pas la panacée qu'on espérait peut-être y trouver. En effet, la survenance se révèle une relation superficielle qui demande qu'on lui trouve un fondement explicatif. Malgré tout, la survenance reste une bonne façon d'identifier ce que soutiennent les physicalistes au sujet du problème corps-esprit.

Au second chapitre, j'examinerai les deux arguments principaux de Kim. Je soutiendrai que le problème de l'exclusion causale/explicative de Kim repose en fait sur le principe de l'exclusion causale plutôt que sur celui de l'exclusion explicative. Je montrerai que seul le principe de l'exclusion causale apparaît bien établi et que le débat entourant le principe de l'exclusion explicative de Kim relève en partie d'un malentendu entre Kim et ses opposants. En effet, les deux partis ne s'entendent pas sur ce qui distingue une explication d'une autre. Je présenterai ensuite en détail l'argument de la survenance, qui est certainement l'argument le plus important de Kim. Cet argument apparemment incontournable réfute le physicalisme non-réductif et explique que Kim se tourne vers le réductionnisme.

Ensuite, au chapitre 3, je présenterai les deux 'solutions' successives qu'a proposées Kim pour expliquer l'attribution de pouvoirs causaux aux états mentaux. On verra que la première proposition de Kim est sujette à des difficultés formelles en plus d'être fondée sur le réductionnisme classique. Quant à la seconde proposition de Kim, elle est particulièrement intéressante car bien qu'elle soit également d'inspiration réductionniste, elle se base principalement sur le fonctionnalisme. La réduction dépendant alors de la possibilité de définir fonctionnellement les états mentaux, il est impossible de réduire les qualia s'ils résistent à la fonctionnalisation[5]. Je discuterai les conséquences de ce fait dans le chapitre 4.

---

[5] Vu la faible quantité de littérature francophone de tradition analytique concernant le problème corps-esprit et suivant l'usage anglais, certains néologismes sont utilisés dans ce mémoire en relation avec la thèse du fonctionnalisme, comme 'fonctionnalisation', 'fonctionnaliser', etc. Quelques autres néologismes du même genre, inspirés de l'anglais, se retrouvent également dans le texte. L'emploi de tels néologismes est destiné, d'une part, à éviter les ambiguïtés que pourrait provoquer l'usage de mots français ayant une signification trop différente de celle du mot anglais et, d'autre part, à faciliter la transition entre ce mémoire et la littérature sur le domaine. Cependant, lorsque des concepts ont reçu une traduction largement diffusée et généralement acceptée, comme par exemple 'survenance' pour l'anglais '*supervenience*', j'emploirai cette traduction.

Le chapitre 4 sera le lieu d'une discussion de certains problèmes liés à l'argumentation de Kim. Par 'problèmes' je n'entends pas principalement des défauts dans le raisonnement, mais plutôt des conclusions indésirables auxquelles est contraint Kim. Pour commencer, je montrerai que Kim exagère la vertu explicative de la réduction. En effet, l'un des avantages les plus importants que Kim voit dans la réduction est le pouvoir explicatif que présente la thèse de l'identité. Je soutiendrai que ce pouvoir explicatif est en fait subjectif et qu'il repose essentiellement sur les intuitions personnelles de chacun. Je discuterai également du problème de l'attribution des pouvoirs causaux dans le cadre du modèle de la réduction fonctionnelle, de même que de l'incapacité de ce modèle à rendre compte des qualia. Enfin, j'examinerai les principales thèses de l'*émergentisme*, notion à laquelle Kim place ses espoirs pour rendre compte des qualia et selon laquelle les propriétés mentales 'émergent' de certaines conditions de base. J'examinerai plus particulièrement une version moderne de l'émergentisme proposé par Humpreys. J'argumenterai que l'émergentisme n'est pas une solution et que cette thèse risque de remettre en cause la stabilité de la position de Kim quant aux états mentaux intentionnels. En effet, on peut prendre l'émergentisme comme impliquant deux choses différentes: soit on considère que l'émergentisme ne respecte pas la fermeture du monde physique, soit il respecte la fermeture du monde physique et ne semble alors pas être en mesure de sauvegarder le mental.

Finalement, en conclusion, je résumerai le chemin parcouru et tirerai les leçons de ce mémoire. Plus particulièrement, j'examinerai si Kim réussit à faire du réductionnisme une option intéressante dans le cadre du problème de la causalité mentale.

Chapitre 1: Notions préliminaires

Avant de pouvoir entrer dans le vif du sujet, il est nécessaire de présenter certaines notions fondamentales qui seront utilisées dans ce mémoire. Quatre notions retiendront notre attention dans ce chapitre. En premier lieu viendra la thèse de la fermeture causale du monde physique. L'acceptation de cette thèse représente une caractéristique importante des débats métaphysiques contemporains. En second lieu, je présenterai le consensus du physicalisme non-réductif. Puisque le physicalisme non-réductif constitue la cible des attaques de Kim, il est important de voir ce qui caractérise ce courant avant de passer aux arguments que Kim lui oppose. J'en profiterai pour donner une brève caractérisation des principales approches que l'on retrouve présentement dans le débat sur la causalité mentale, en les regroupant en deux catégories selon qu'elles supportent ou s'opposent au physicalisme non-réductif. En troisième lieu, nous verrons la notion de quale. J'ai déjà mentionné et présenté rapidement les qualia dans l'introduction, mais une discussion un peu plus élaborée n'est pas superflue étant donné que les qualia sont les états mentaux qui donnent le plus de fil à retordre aux philosophes. Cette même raison explique également pourquoi les qualia joueront un rôle important dans les discussions de ce mémoire. Enfin, je discuterai de la survenance, une notion importante dans les discussions sur le problème corps-esprit et qui joue un rôle de premier plan dans l'un des arguments principaux de Kim, l'argument de la survenance.

La fermeture causale du monde physique

Les réalisations impressionnantes des sciences de la nature nous portent à croire que nous serons un jour capables d'expliquer la plupart sinon tous les phénomènes se présentant à nous à l'aide de la science. Cette conviction (ou cet espoir) nous conduit à poser que tous les phénomènes physiques sont explicables physiquement ('physique' et 'physiquement' sont ici et dans ce qui suit pris dans un sens large, c'est-à-dire incluant non seulement la physique proprement dite mais également la chimie et la biologie). Il s'agit là de la thèse de la *fermeture causale du monde physique*: seuls les événements physiques peuvent causer des événements physiques. Cette thèse est supportée par la thèse de la *détermination causale du monde*

physique: pour n'importe quel événement physique (causé) *e*, il y a une chaîne composée uniquement d'événements physiques conduisant à *e* dont chaque maillon détermine causalement le suivant (ou plutôt, étant donné que le déterminisme strict n'apparaît pas défendable à la lumière des dernières avancées de la physique, chaque maillon détermine la probabilité objective d'apparition du maillon suivant)[6]. Non seulement ces thèses découlent-elles, d'une certaine manière, des sciences de la nature, mais elles en constituent également des postulats nécessaires. En effet, grâce à ces thèses, lorsque les scientifiques sont confrontés à un phénomène physique à expliquer, ils n'ont pas à se questionner sur la pertinence de leur recherche d'une cause physique car en aucun cas, par principe, la cause recherchée se révélera-t-elle non-physique.

Le consensus du "physicalisme non-réductif"

Qu'en est-il du mental alors? Une approche *matérialiste*, c'est-à-dire respectant la fermeture causale du monde physique, considère le mental comme un cas spécial de phénomène physique. Dire que le mental est un cas spécial de phénomène physique, c'est rejeter l'existence d'une substance différente pour le mental: je l'ai dit, le dualisme des substances à la Descartes est aujourd'hui largement désavoué. Les philosophes contemporains sont très largement matérialistes ou, ce qui est *grosso modo* la même chose, *physicalistes*. Ce physicalisme se caractérise par la conviction qu'il n'existe qu'une seule substance plutôt que deux et que cette substance est *physique*. Toutefois, ce monisme des substances ne conduit pas nécessairement à l'élimination ou à la réduction du mental au physique. En effet, à ce monisme de substance s'allie chez une majorité de philosophes un *dualisme des propriétés*, c'est-à-dire qu'il existe, selon eux, à la fois des propriétés physiques et des propriétés mentales. Un phénomène mental est un cas spécial de phénomène physique en ce qu'il partage la même substance physique mais possède des propriétés particulières, les propriétés mentales. Dans ce contexte, la priorité du physique s'exprime par le fait que s'il peut exister des objets n'ayant

---

[6] Ces définitions sont celles de McLaughlin. "*Physical Causal Closure: Only physical events can cause physical events.*" "*Physical Causal Determination: For any (caused) physical event, P, there is a chain of entirely physical events leading to P, each link of which causally determines its successor.*" *in* Guttenplan (1994) p. 281.

que des propriétés physiques (par exemple une chaise, un rocher, etc.), de même que des objets ayant à la fois des propriétés physiques et mentales (les êtres humains, bien sûr, mais aussi probablement certains animaux supérieurs et d'éventuels extraterrestres), il ne peut pas exister d'objets possédant uniquement des propriétés mentales (comme des "fantômes", des "dieux" ou autres "purs esprits"). Une dernière caractéristique de ce dualisme des propriétés est que les propriétés mentales ne peuvent être ramenées ou réduites à des propriétés physiques; autrement il n'y aurait plus deux types fondamentaux de propriétés, mais bien un seul.

À ce dualisme des propriétés s'ajoute la thèse du réalisme du mental. Le réalisme du mental implique deux choses: d'une part, que les propriétés mentales existent, ce qui est déjà impliqué par le dualisme des propriétés et, d'autre part, qu'un événement puisse manifester des pouvoirs causaux *en vertu du fait qu'il possède une propriété mentale*. Ainsi, les événements mentaux, en tant qu'ils possèdent des propriétés mentales, sont réputés pouvoir interagir causalement avec les autres événements mentaux et les événements physiques. Cette insistance sur le fait que l'interaction causale se fait en vertu des propriétés mentales d'un événement est fondamentale et est destinée à répondre à deux problèmes. D'une part, il s'agit de préciser la manière dont s'effectue la relation de causalité en question. En effet, comme on l'a vu, le physicalisme non-réductif se prononce *contre* l'existence d'événements ne possédant que des propriétés mentales. Par conséquent, tout événement ayant des propriétés mentales doit également posséder des propriétés physiques. En présence d'un tel événement *a* qui causerait un autre événement *b* (ayant ou non des propriétés mentales en plus de ses propriétés physiques), on peut donc légitimement se demander si l'interaction causale entre *a* et *b* se fait en vertu des propriétés physiques ou des propriétés mentales de *a*. En insistant sur le fait que l'interaction causale se fait en vertu des propriétés mentales de *a*, on répond à cette question. Une telle affirmation ne constitue cependant pas une justification. Cette justification est beaucoup plus difficile à obtenir et constitue en fait l'essence même du problème de la causalité mentale.

D'autre part, en spécifiant que l'interaction causale se fait en vertu des propriétés mentales de *a*, on empêche que les pouvoirs causaux de *a* soient entièrement attribués aux propriétés

physiques de *a*, rendant les propriétés mentales inefficaces. En effet, puisque l'attribution de pouvoirs causaux à des propriétés physiques est beaucoup moins problématique que l'attribution de pouvoirs causaux à des propriétés mentales, il peut être tentant de dire que les événements mentaux possèdent bien des pouvoirs causaux, mais en vertu du fait que tout événement mental est également un événement physique. Cette approche conduit cependant à l'épiphénoménisme, la thèse selon laquelle les événements mentaux existent bel et bien, mais n'ont aucun rôle causal à jouer. Cependant, selon un certain point de vue, l'épiphénoménisme peut être considéré comme une forme d'éliminativisme, la thèse selon laquelle les événements mentaux n'existent pas[7]. Plusieurs auteurs partagent ce point de vue, y compris Kim. Selon ce point de vue, il ne fait aucun sens de parler de l'existence d'une chose qui n'aurait pas de pouvoirs causaux: "Car un critère plausible pour distinguer ce qui est réel de ce qui ne l'est pas est la possession de pouvoirs causaux. Comme l'a dit Samuel Alexander, quelque chose qui "ne sert à rien, n'a pas de fonction à remplir" - c'est donc dire quelque chose n'ayant pas de pouvoirs causaux - "peut aussi bien être aboli, et sans aucun doute le sera en temps et lieu."[8] Contre cette approche, on peut remarquer que l'existence d'une chose et la détermination de ses pouvoirs causaux constituent deux questions différentes. D'un autre côté, on peut se demander comment nous pourrions en venir à constater l'existence de quelque chose n'ayant aucun impact sur le monde. Dire que les événements mentaux n'ont pas de pouvoirs causaux revient donc, d'une certaine façon, à dire qu'ils n'existent pas. La thèse du réalisme du mental est par conséquent directement opposée à l'épiphénoménisme, la thèse selon laquelle les états mentaux existent, mais n'ont aucun pouvoir causal. Le réalisme du mental fait ainsi un pas de plus que le dualisme des propriétés qui, lui, est compatible avec l'épiphénoménisme.

---

[7] Les définitions sommaires de l'épiphénoménalisme et de l'éliminativisme données ici n'entendent pas rendre justice à la pensée des partisans de ces approches, mais seulement fournir une caractérisation qui permet de les différencier et qui correspond *grosso modo* à la discussion de ces notions présentée plus loin.

[8] "*For a plausible criterion for distinguishing what is real from what is not real is the possession of causal power. As Samuel Alexander said, something that "has nothing to do, no purpose to serve" - that is, something with no causal power - "might as well, and undoubtedly would in time, be abolished."* " Kim (1998), p. 119. La citation de Samuel Alexander provient de *Space, Time, and Deity* (London, Macmillan, 1927), vol. 2, p. 8.

Ce monisme des substances avec dualisme des propriétés, allié à la thèse du réalisme du mental, représente aujourd'hui un large consensus et constitue la position métaphysique la plus influente quant au problème corps-esprit: *le physicalisme non-réductif.* Ce consensus n'entraîne cependant pas la création d'un réel paradigme. En effet, le réalisme du mental ne précise pas *la manière* dont un événement peut en causer un autre en vertu de ses propriétés mentales. C'est pourquoi la question se pose de savoir si le réalisme du mental est réellement compatible avec la fermeture causale du monde physique. Cet état de fait a conduit à l'apparition de nombreuses approches concurrentes tentant de répondre à cette question du *'comment'*, dont les deux principales sont le monisme anomal (Donald Davidson) et le fonctionnalisme (du moins toutes ses variantes non-dualistes: Hilary Putnam, David M. Armstrong, Jerry Fodor, Sydney Shoemaker, etc.). Le physicalisme non-réductif est donc mieux compris comme le cadre dans lequel se déroule les débats, plutôt que comme une stratégie de recherche particulière. Outre le dualisme des substances (Descartes), sont *exclus* de ce consensus: l'éliminativisme (Paul M. Churchland, Patricia S. Churchland et, en ce qui concerne les qualia, Daniel C. Dennett), le réductionnisme (Smart et Feigl) et l'épiphénoménisme (Joseph Levine). Comme on peut le constater, en soutenant que nous n'avons d'autres options que le réductionnisme ou l'épiphénoménisme, Kim s'oppose directement aux thèses du physicalisme non-réductif.

Principales approches supportant le physicalisme non-réductif

Le monisme anomal

Le monisme anomal est l'approche défendue par Davidson et présentée principalement dans son article "Mental Events" (1970). Cette approche a eu beaucoup d'adeptes à un certain moment et est encore défendue avec acharnement par plusieurs philosophes. L'attrait du monisme anomal vient de ce qu'il correspond bien au physicalisme non-réductif. Le *monisme* du monisme anomal est celui *de substance*: conformément au physicalisme non-réductif, le monisme anomal considère que la seule sorte de substance qui existe est physique. Davidson ne s'exprime pas en termes de propriétés; on ne peut donc pas dire qu'il supporte le dualisme des propriétés. Malgré tout, Davidson croit en l'existence des phénomènes mentaux et en

l'impossibilité de les réduire au physique. D'après lui, la connaissance complète du cerveau et de ses processus ne nous donnerait aucune connaissance sur le mental, c'est-à-dire sur les croyances, les désirs, les expériences phénoménales, etc. La raison en est que le mental et le physique sont régis par des systèmes de normes différents et irréductibles l'un à l'autre. Il en découle une impossibilité de formuler des lois reliant le mental au physique. L'expression 'anomalisme du mental' est utilisée pour exprimer le fait que le mental ne tombe pas sous des lois.

Davidson soutient également que les événements mentaux sont la *cause* des actions dont on tient les agents communément responsables. Davidson explique cette capacité causale des événements mentaux en identifiant les différentes *occurrences* d'événements mentaux avec des événements physiques. Par contre, il est faux de dire que les *types* d'événements mentaux sont identifiés avec des types d'événements physiques, comme le soutient le réductionnisme. Qu'est-ce que cela veut dire? Cela signifie que malgré qu'il soit *faux*, par exemple, de dire que *toute douleur* causée par une brûlure correspond à l'activation des fibres-C, il peut être *vrai* de dire que telle *occurrence* de douleur, par exemple celle causée par la brûlure que je me suis faite à la main hier soir, correspond à l'activation de fibres-X. Par contre, la sensation de douleur causée par une brûlure que je me suis faite à la même main mais à un autre moment correspondait possiblement à un tout autre phénomène physique que l'activation de fibres-X ou de fibres-C. Bref, un type de phénomènes mentaux n'est pas associé avec un type de phénomènes physiques, mais toute occurrence d'un phénomène mental doit être accompagné de l'occurrence d'un phénomène physique quelconque. Cet *'accompagnement'* des phénomènes mentaux par des phénomènes physiques est désigné par Davidson comme étant une relation de *survenance*, sur laquelle je reviendrai plus loin. L'important pour le monisme anomal, c'est qu'il est impossible d'établir des lois entre les types de phénomènes mentaux et les types de phénomènes physiques, autrement l'anomalisme du mental serait rompu.

J'ai dit que le monisme anomal correspondait bien au physicalisme non-réductif. La vérité est que le monisme anomal est *antérieur* au consensus du physicalisme non-réductif, aussi serait-il

plus juste de dire que le monisme anomal, avec le fonctionnalisme, est à l'origine de ce consensus.

Le fonctionnalisme

Le fonctionnalisme, pris au sens large, représente la position dominante aujourd'hui face au problème corps-esprit. Dans "What is Functionalism?" (1980), Ned Block distingue deux courants majeurs se retrouvant sous la bannière du fonctionnalisme: *l'analyse fonctionnelle* et *le fonctionnalisme métaphysique*. Le courant de l'analyse fonctionnelle s'intéresse à la recherche *d'explication fonctionnelle*, c'est-à-dire des explications basées sur la décomposition d'un système en parties plus simples, le système étant expliqué par les caractéristiques des parties et la façon dont elles interagissent entre elles. Par exemple, on explique comment l'automobile se déplace en montrant comment le moteur produit du mouvement qui, par le biais de la transmission puis de l'essieu, est finalement transmis aux roues; le fonctionnement du moteur, à son tour, peut être expliqué de façon similaire. Une application particulière de cette idée de l'analyse fonctionnelle est la comparaison de notre esprit avec un programme d'ordinateur: une analyse fonctionnelle de nos processus mentaux révélerait que notre esprit se ramène à des opérations aussi simples que celles d'un ordinateur digital.

Pour sa part, le fonctionnalisme métaphysique se prononce sur la *nature* de l'esprit, plutôt que de seulement proposer une stratégie de recherche d'explication. En effet, pour les fonctionnalistes, les états mentaux ne sont rien d'autre que des *états fonctionnels*. Ces états fonctionnels sont définis par leurs antécédents causaux (des entrées sensorielles et/ou d'autres états mentaux) et leurs produits causaux (des comportements et/ou d'autres états mentaux). Le fonctionnalisme métaphysique est souvent perçu comme une simple reformulation plus sophistiquée du béhaviorisme. Il est vrai que le fonctionnalisme se prononce sur le statut des états mentaux et utilise les stimuli et les réponses comme base de sa caractérisation des différents 'états'. Mais alors que le béhaviorisme éliminait les états mentaux pour ne parler que de simples *dispositions* ne devant pas être interprétées comme indiquant la présence d'états internes, le fonctionnalisme, lui, admet l'existence d'états internes ayant un rôle causal à jouer. C'est pourquoi la caractérisation d'un état fonctionnel inclut la référence à d'autres états

fonctionnels en plus des stimuli et des réponses. Il faut également noter que cette identification des états mentaux avec des états fonctionnels se fait au plan des *types* et non pas seulement au plan des occurrences. Pour un fonctionnaliste, ce que les diverses occurrences de douleur ont en commun, c'est un rôle causal. Toutefois, pour chaque occurrence, ce rôle causal peut être réalisé par une variété de systèmes. La grande majorité des fonctionnalistes étant physicalistes, le système qui *réalise* telle occurrence x d'un état fonctionnel F est considéré comme physique (s'il ne l'est pas, il s'agit alors d'une forme de fonctionnalisme qui ne peut pas être inclus sous la bannière du physicalisme non-réductif). Le fonctionnalisme permet la *réalisation multiple*, c'est-à-dire qu'un état mental M, en tant qu'il consiste en un état fonctionnel, peut être réalisé par différents mécanismes selon les espèces ou même à l'intérieur d'une même espèce, voire chez un même individu (à cause de transformations neurologiques au cours de la croissance ou suite à un accident, par exemple). Il est compatible avec le fonctionnalisme d'attribuer à des extraterrestres ou à des robots sophistiqués les mêmes états mentaux qu'aux humains, puisqu'il suffit que les mêmes états fonctionnels soient présents. Les matériaux constitutifs et l'organisation particulière de ces états fonctionnels n'ont pas d'importance fondamentale.

Principales approches s'opposant au physicalisme non-réductif

Le réductionnisme

De façon générale, le réductionnisme est une approche explicative qui consiste à ramener un phénomène à d'autres déjà connus. L'exemple classique concerne la température: la température d'un corps s'est révélée n'être rien d'autre que l'énergie cinétique moyenne des molécules de ce corps. En identifiant un phénomène macroscopique (la température) avec un phénomène microscopique (le mouvement des molécules), nous n'avons plus deux phénomènes à expliquer, mais un seul: le mouvement des molécules, qui est un phénomène beaucoup plus facile à expliquer que la température en elle-même. Le réductionnisme implique *une identité des types*: un type de phénomènes (la température) est identifié à un autre type de phénomènes (l'énergie cinétique des molécules). Ce n'est pas seulement que *telle occurrence* de température est identifiée avec *telle occurrence* de mouvement des molécules, mais que *toute occurrence* de température peut être ramenée à une occurrence de mouvement des

molécules. Suivant Ernest Nagel (1961), on parle généralement de *réduction interthéorique*. En effet, dans l'exemple de la température, une théorie, la thermodynamique classique, se trouve ramenée ou réduite à une autre théorie, la mécanique statistique moléculaire. Pour effectuer une telle réduction, il faut non seulement posséder une théorie du phénomène macroscopique et une théorie du phénomène microscopique, mais également des 'lois-ponts' (*bridge-laws*), c'est-à-dire des lois établies empiriquement qui nous montrent que le phénomène macroscopique est coextensif avec le phénomène microscopique. Ces 'lois-ponts' nous permettent de dériver la théorie macroscopique de la théorie microscopique, effectuant ainsi la réduction.

Appliqué au problème corps-esprit, le réductionnisme, ou *théorie de l'identité des types*, consiste simplement à soutenir que les différents types d'états mentaux sont identiques à des états physiques, généralement des états neurologiques du cerveau (Smart, 1959). Les processus mentaux n'étant "rien d'autre" que les processus de notre cerveau, il suffit d'expliquer ces derniers pour rendre compte de nos états mentaux. Le réductionnisme offre l'avantage d'être pleinement en accord avec le matérialisme. Toutefois, pour plusieurs philosophes, le réductionnisme est trop exigeant. Premièrement, il y a des problèmes pratiques: nous n'avons pas encore de neurologie complète pouvant servir de théorie microscopique, ni de psychologie complète pouvant servir de théorie macroscopique et donc encore moins de 'lois-ponts'. Mais ces problèmes pratiques ne sont justement que ça: des problèmes pratiques qui peuvent être résolus; c'est là le travail des scientifiques, pas des philosophes. Deuxièmement, cependant, se trouvent des problèmes conceptuels: plusieurs philosophes soutiennent que les qualia, cet aspect subjectif de nos perceptions et sensations, ne peuvent être réduits à des propriétés physiques. Puisqu'une connaissance exhaustive des processus et des structures du cerveau ne pourra jamais nous donner accès aux qualia, les qualia ne peuvent être identiques à ces processus. En effet, deux objets identiques doivent nécessairement posséder toutes les mêmes propriétés. Troisièmement, l'argument de l'anomalisme du mental de Davidson contrevient également à l'identité des types: si on ne peut établir de lois entre le mental et le physique, alors on ne peut pas obtenir de 'lois-ponts'. De plus, si le mental était identique au physique, il y aurait nécessairement certaines régularités (ne serait-ce que des régularités d'occurences) entre

les deux types de phénomènes. L'absence de telles régularités indique la présence de deux objets, propriétés, événements, etc. différents. Enfin, le réductionnisme ne permet pas la *réalisation multiple*. Si les processus mentaux sont identifiés avec des processus du cerveau humain, cela voudrait dire que nous éliminons *a priori* la possibilité que des êtres ne possédant pas ces mêmes structures puissent néanmoins avoir des états mentaux. Or, il nous semble déjà probable que les animaux (du moins les mammifères) fassent l'expérience de certains états mentaux, en particulier les qualia, et nous ne pouvons écarter la possibilité de rencontrer un jour des extraterrestres qui, bien que possédant des structures biologiques radicalement différentes des nôtres, nous apparaissent néanmoins comme possédant des états mentaux semblables aux nôtres.

L'éliminativisme

Les éliminativistes comme Dennett et les Churchland considèrent que la conception ordinaire du mental (c'est-à-dire comme comprenant des croyances, des désirs, des expériences phénoménales, etc.) comporte des difficultés conceptuelles majeures. Cette conception du mental, souvent appelée *'psychologie populaire'* (*folk psychology*), est d'après les éliminativistes tellement inappropriée qu'il s'avère impossible de la réduire aux sciences de la nature. Or, sans réduction, disent les éliminativistes, le mental devient un domaine complètement atypique n'offrant ni perspective d'explication satisfaisante par la psychologie populaire, ni intégration satisfaisante avec le reste de nos théories scientifiques. C'est pourquoi les éliminativistes proposent *d'éliminer* la conception ordinaire du mental pour la remplacer par une autre qui s'accorde mieux avec les sciences de la nature. Ainsi, le problème corps-esprit ne viendrait pas du fait que l'esprit possède des caractéristiques uniques, mais bien du fait que nous entretenons des idées erronées à son sujet. Une meilleure conception du mental nous permettrait d'incorporer le mental dans le même cadre ontologique et épistémologique que le reste des objets et événements de l'univers spatio-temporel. Ainsi, suivant la thèse de l'éliminativisme, le mental *per se* n'est pas vraiment éliminé; seule la conception ordinaire du mental (partagée par la plupart des philosophes) l'est. Reste qu'éliminer cette conception ordinaire revient à nier l'existence de ce 'mental' auquel réfèrent sans cesse les philosophes

non-éliminativistes. C'est seulement de ce 'mental traditionnel' dont on parle lorsque l'on résume la position des éliminativistes en disant qu'ils nient l'existence du mental.

L'épiphénoménisme

Bien qu'il nous soit tout à fait naturel de référer à nos états mentaux pour expliquer nos actions et celles des autres (Marie prend son parapluie parce qu'elle croit qu'il va pleuvoir, Jean crie parce qu'il ressent une sensation de brûlure, etc.), les épiphénoménistes croient qu'il s'agit là d'une erreur, que l'apparence est trompeuse. La thèse de l'épiphénoménisme est qu'en réalité les phénomènes mentaux n'ont aucun pouvoir causal, que les phénomènes mentaux sont en soi causalement inefficaces. Seuls les phénomènes physiques peuvent exercer des pouvoirs causaux. Des phénomènes physiques sont en fait responsables de chacune de nos actions, *de même que de l'existence des phénomènes mentaux eux-mêmes*. Si les phénomènes mentaux accompagnent de manière si régulière nos actions, c'est pour une raison bien simple: nos phénomènes mentaux et nos actions sont tous deux les produits d'une cause commune, certains phénomènes physiques (ou, plus précisément, biologiques) qui ont lieu dans notre corps.

Le grand avantage de l'épiphénoménisme est d'accepter l'existence des phénomènes mentaux et d'en expliquer l'origine. L'épiphénoménisme règle aussi le problème corps-esprit d'une certaine manière: il n'y a plus besoin d'expliquer comment les phénomènes mentaux peuvent exercer leurs pouvoirs causaux s'ils n'en possèdent pas. Par contre, si les philosophes désirent tant rendre compte de la manière dont les phénomènes mentaux exercent leurs pouvoirs mentaux, c'est justement parce qu'ils considèrent que les phénomènes mentaux possèdent des pouvoirs causaux. Autrement dit, l'épiphénoménalisme n'est pas une option intéressante pour qui accepte le réalisme du mental. De plus, j'ai déjà mentionné que pour plusieurs philosophes l'épiphénoménisme revient à peu de chose près à de l'éliminativisme. En quoi le concept de quelque chose d'inutile (comme des phénomènes mentaux causalement inefficaces) pourrait-il nous être utile? Puisque selon la thèse de l'épiphénoménisme toute explication se fait sur la base des phénomènes physiques, nous pouvons aussi bien éliminer les phénomènes mentaux de notre conception du monde, ce qui présente l'avantage de simplifier notre ontologie. Évidemment, il ne s'agit pas là de l'opinion des épiphénoménalistes eux-mêmes.

Les qualia

Les nombreuses approches au problème corps-esprit ne traitent pas nécessairement tous les états mentaux de la même manière. En effet, la conscience phénoménale pose des problèmes particuliers qui sont plus ou moins bien résolus par les différentes approches. Pour plusieurs philosophes, les qualia constituent la classe d'états mentaux la plus difficile à traiter et à expliquer. Leur définition elle-même pose problème, comme j'en ai donné un aperçu dans l'introduction. En effet, la conscience phénoménale est conçue comme distincte des notions de cognition, de traitement de l'information ou de fonctionnalisation. Les qualia apparaissent n'avoir aucun effet observable de façon objective ou même intersubjective. C'est que les qualia ne sont pas accessibles à un observateur externe: ils ne sont accessibles qu'à la personne qui les ressent. De plus, ils nous apparaissent intrinsèquement distincts de tous les autres phénomènes du monde; il semble intuitivement impossible que nous puissions les expliquer de la même manière que les phénomènes physiques qu'étudie la science. Trois 'arguments' ont été élaborés pour tenter de rendre compte du caractère particulier de la conscience phénoménale: *l'argument du fossé explicatif* (*the explanatory gap argument*), *l'argument de la connaissance* (*the knowledge argument*) et enfin les arguments de *l'hypothèse du spectre inversé* et des *qualia absents*, qui reposent sur la même intuition.

Trois arguments en faveur des qualia

1- L'argument du fossé explicatif

L'argument du fossé explicatif, dû à Levine (1983), insiste sur le fait que même si le réductionnisme était vrai, nous ne pourrions pas vraiment rendre compte du lien entre les qualia et leurs bases physiologiques. Supposons par exemple que la sensation de douleur soit identifiée avec l'activation de fibres-C. Levine argumente qu'il reste impossible d'expliquer pourquoi la douleur produit l'effet qu'elle fait, ou pourquoi cette sensation particulière est liée à cette base particulière. Le lien entre les qualia et leurs bases physiologiques serait totalement arbitraire: il s'agit d'un fait brut que nous devons accepter. Il y a donc un fossé entre ce que nous pouvons expliquer et ce que nous voudrions pouvoir expliquer des qualia. Certains

philosophes considèrent ce fossé comme étant insurmontable pour des raisons de principes liées à la nature des qualia eux-mêmes (comme par exemple Levine et McGinn) alors que d'autres concluent plus drastiquement à l'inexistence des qualia (Dennett). Il demeure cependant possible que notre incapacité à combler le fossé explicatif soit simplement due au fait qu'il nous manque les concepts nécessaires. C'est ce que suggère une analogie de Nagel (1974) selon laquelle le lien entre les qualia et notre cerveau est aussi incompréhensible pour nous que l'équivalence entre la masse et l'énergie peut l'être à un homme des cavernes. Bien qu'extrêmement difficile, le problème corps-esprit pourrait ainsi éventuellement être résolu.

2- L'argument de la connaissance

L'argument de la connaissance est plus ambitieux que l'argument du fossé explicatif: il vise à montrer qu'il est impossible de ramener les qualia à quelque chose de physique. D'après cet argument, il existe des connaissances qui ne peuvent être acquises qu'à la condition de soi-même vivre l'expérience phénoménale. Le cas le plus célèbre d'application de l'argument de la connaissance est l'exemple de Jackson (1982, 1986): une hypothétique super-spécialiste de la vision, Mary, a passé toute sa vie dans un environnement en noir et blanc. Malgré tout, grâce à la télévision (en noir et blanc, bien sûr), Mary est devenue une experte mondiale de la vision couleur et, par hypothèse, elle connaît tout ce qu'il y a à connaître de physique à propos de la perception visuelle du rouge par un sujet normal. Jackson soutient alors que Mary ne sait pas absolument tout ce qu'il y a à savoir sur la perception du rouge puisqu'elle-même n'en a jamais fait l'expérience: si Mary sortait de son environnement monochromatique et percevait pour la première fois du rouge, elle apprendrait alors quelque chose de nouveau, elle apprendrait "ce que ça fait" que de voir du rouge. Mais comme Mary, par hypothèse, connaît tout ce qu'il y a à savoir de physique sur la vision des couleurs, il faut conclure que les qualia ne sont pas physiques.

Cependant, bien que la plupart des philosophes concèdent à Jackson que Mary apprend bien quelque chose lorsqu'elle voit du rouge pour la première fois, de nombreuses questions quant à *la nature* de ce qu'elle apprend ont été soulevées. Lawrence Nemirow (1980) et David Lewis (1983) soutiennent que la seule chose que Mary acquiert est un savoir-faire, c'est-à-dire une

nouvelle habileté pratique à reconnaître et à imaginer les qualia, et plus particulièrement le quale de rouge. Par conséquent, en sachant tout ce qu'il y a à savoir de physique sur la vision, Mary sait vraiment tout ce qu'il y a à savoir sur ce sujet. Toutefois, s'il est vrai que Mary apprend bien une nouvelle habileté, il est moins clair qu'il s'agit là de la seule chose qu'elle apprend en voyant du rouge pour la première fois. Intuitivement, il semble bien que Mary acquiert de nouvelles connaissances, pas seulement de nouvelles habiletés. Reste à savoir de quel genre de connaissances il s'agit. P. M. Churchland (1985) et Michael Tye (1986) soutiennent que ce que Mary gagne lorsqu'elle perçoit du rouge pour la première fois est *une nouvelle manière d'accéder* à une connaissance qu'elle possède déjà par ailleurs. Cette nouvelle façon peut se caractériser par le fait que Mary connaît maintenant directement et par introspection ce qu'elle ne connaissait avant qu'indirectement par inférence. Ou encore par le fait que Mary peut maintenant se représenter les qualia d'une manière pré-linguistique sans avoir à faire appel à ses capacités de raisonnement. Reste qu'encore une fois la conclusion est qu'en sachant tout ce qu'il y a à savoir de physique sur la vision, Mary sait vraiment tout ce qu'il y a à savoir sur ce sujet. Or, pour que l'argument de Jackson soit valide, il faut nécessairement que Mary acquiert de nouvelles *connaissances*. Au bout du compte, le pouvoir de persuasion de l'argument de la connaissance dépend grandement de nos propres intuitions: est-ce que savoir "ce que cela fait" que de voir du rouge pour la première fois constitue une véritable acquisition de connaissances ou est-ce qu'au contraire cela se limite à l'apprentissage d'une nouvelle habileté ou d'une nouvelle manière d'accèder à des connaissances qui peuvent être acquises autrement?

3- Les arguments du spectre inversé et des qualia absents

L'argument du spectre inversé et celui des qualia absents, quant à eux, reposent sur l'intuition que les qualia ne possèdent pas de lien nécessaire avec leur base physiologique, qu'ils ne semblent qu'arbitrairement liés à elle. L'argument du spectre inversé repose sur le fait que les qualia sont nécessairement perçus à la première personne: je n'ai aucune garantie que ce que *je* perçois comme du rouge correspond vraiment à ce que *quelqu'un d'autre* perçoit comme du rouge: tout ce que *je* sais, c'est que nous nous entendons pour dire qu'une tomate mûre ou un

nez de clown sont rouges. L'argument du spectre inversé se base sur ce genre de raisonnement pour montrer qu'il est tout à fait possible qu'un individu perçoive le rouge comme je perçois le vert, le vert comme je perçois le rouge, et ainsi de suite pour toutes les couleurs. À la vision d'un objet d'une certaine couleur pour moi, cet individu percevrait le quale inversé du mien. Cependant, tout comme j'ai appris à appeler "rouge" la couleur d'une tomate mûre et "verte" la couleur de la pelouse, cet individu emploierait le mot "rouge" lorsqu'il pointe vers une tomate mûre et "verte" lorsqu'il pointe vers la pelouse. Le comportement de cet individu serait par conséquent semblable au mien à tous les égards. Il pourrait passer toute sa vie avec un spectre inversé par rapport au mien sans que cela ne fasse la moindre différence fonctionnelle. En fait, d'après l'argument du spectre inversé, rien ne garantit que ce ne soit d'ailleurs le cas: peut-être que dans notre monde les gens n'ont pas tous les mêmes qualia pour les mêmes objets. L'argument des qualia absents va dans le même sens. Posons-nous la question: à quoi servent les qualia? Il ne semble pas y avoir de fonction associée de façon nécessaire aux qualia. Pour distinguer le feu vert du feu rouge à une intersection, je n'ai pas besoin de conscience phénoménale: tout ce qu'il me faut c'est de l'*information* sur l'état présent du feu de circulation. Or, la conscience phénoménale est radicalement distincte de la notion d'information. C'est pourquoi on peut construire des machines capables de distinguer entre un feu vert et un feu rouge sans qu'on leur suppose de conscience phénoménale. Ne pourrait-on pas alors imaginer un monde fonctionnellement identique au nôtre, c'est-à-dire où tout le monde réagirait exactement de la même manière que nous et utiliserait même des mots tels que "douleur", "rouge", "salé", etc., mais dans lequel tous les habitants seraient des "zombies" phénoménaux, c'est-à-dire qu'ils n'auraient aucune conscience phénoménale? C'est là l'hypothèse des qualia absents. On peut toutefois s'interroger: si un monde fonctionnellement identique au nôtre mais sans qualia est possible, c'est dire que les qualia ne jouent aucun rôle dans notre survie, qu'il s'agit de simples épiphénomènes. Mais dans ce cas, pourquoi l'évolution aurait-elle sélectionné la conscience phénoménale dans notre monde? Et s'il s'agit d'épiphénomènes, comment en sommes-nous venus à autant parler des qualia? Ne sont-ils pas supposés n'avoir aucun impact sur notre vie? L'idée des zombies phénoménaux est logiquement consistante (semble-t-il) mais n'en est pas moins peu plausible de l'avis général.

Quant à l'argument des qualia inversés, ce que nous savons de nos organes sensoriels rend impossible de telles inversions sans que des différences fonctionnelles (parfois mineures, mais différences tout de même) apparaissent et ce, non seulement pour la vue mais également pour les autres sens. Reste que ces arguments, de même que l'argument du fossé explicatif et l'argument de la connaissance, nous aident à mieux comprendre la notion de quale et permettent de préciser à quelles questions devrait être en mesure de répondre un modèle proposant une explication de *tous* les états mentaux, incluant la conscience phénoménale.

Le cas de la survenance

La survenance a été souvent perçue comme la solution au problème corps-esprit par les partisans du physicalisme non-réductif. Il s'agit d'un type de relation pouvant exister entre deux propriétés ou caractéristiques d'un événement, comme l'est la corrélation ou la causalité. Le sens particulier qu'acquiert cette notion de survenance en philosophie de l'esprit provient du célèbre article "Mental Events" de Davidson. Dans cet article, Davidson mentionne la survenance lorsqu'il décrit sa propre théorie, le monisme anomal. Voici ce qu'il dit de la survenance:

> Bien que la position que je décrive nie l'existence de lois psychophysiques, elle est compatible avec l'idée que les caractéristiques mentales sont d'une certaine manière dépendantes ou survenantes par rapport aux caractéristiques physiques. Une telle survenance doit être prise comme signifiant que deux événements en tout point semblables physiquement ne peuvent différer mentalement de quelque façon que ce soit, ou alors qu'un objet ne peut changer mentalement sans présenter une variation physique quelconque. Ce genre de dépendance ou de survenance n'implique pas la réductibilité par le biais de loi ou de définition[9]

La survenance assure qu'il ne peut y avoir présence de propriétés mentales sans que soient également présentes des propriétés physiques. Par contre, le type de dépendance impliqué par

---

[9] "*Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect. Dependence or supervenience of this kind does not entail reducibility through law or definition*" 1970, p. 214.

la survenance reste peu clair. En effet, ce passage de Davidson présente en fait *deux* caractérisations de la relation de dépendance qui ont chacune des implications différentes. D'une part, Davidson déclare qu'il ne peut y avoir deux événements semblables physiquement qui différeraient mentalement. Cette formulation n'exclut pas la possibilité que deux événements *dissemblables* physiquement soient semblables mentalement. Mais d'autre part, Davidson nous dit également qu'un objet ne peut changer mentalement sans changer physiquement. Cette seconde formulation semble permettre qu'un objet ayant une propriété mentale M_1 survenant sur une propriété physique P_1 à un temps t_1 puisse changer une première fois pour se retrouver au temps t_2 avec une propriété mentale M_2 survenant sur une propriété physique P_2, puis changer à nouveau à t_3 pour *revenir* à la propriété physique P_1 mais sur laquelle survient cette fois M_3. Dans ce dernier cas, la survenance de M_3 sur P_1 à t_3 apparaît comme une violation de la première formulation puisque M_1 aussi peut survenir sur P_1 à t_1. Deux événements semblables physiquement diffèrent donc mentalement. Et même si nous limitions la seconde formulation de Davidson pour contrecarrer l'exemple que je viens de donner en disant que deux événements *simultanés* ne peuvent être semblables physiquement et différer mentalement, il reste que si un même objet, à deux moments différents, peut contrevenir à ce principe, alors il est plausible de croire que deux objets différents peuvent le violer au même moment.

Ce que nous apprend cette petite discussion, c'est que le concept de survenance se prête à de multiples variantes. Davidson, pour sa part, privilégie sa première formulation à sa seconde. De façon générale, la survenance réfère à une forme de dépendance entre deux propriétés, traits, caractéristiques, etc. Cette dépendance n'a pas la force de la causalité, non plus que celle de l'identité, mais elle représente plus qu'une simple corrélation.

Voici un exemple pour mieux comprendre l'idée générale de la survenance. Lorsque je peins un tableau, je désire produire un objet possédant des propriétés esthétiques. Cependant, les seules variables que je manipule sont physiques: le type, la grandeur et la forme de la toile, le type de peinture utilisée, les couleurs employées et leur disposition sur la toile. Une fois mon tableau peint et ses propriétés physiques déterminées, je n'ai pas à retrousser mes manches et

à m'attaquer ensuite aux propriétés esthétiques de mon tableau; lorsque les propriétés physiques d'un tableau sont fixées, ses propriétés esthétiques le sont aussi. Pourtant, il semble absurde d'identifier les propriétés esthétiques d'un tableau à la peinture qui le recouvre. La beauté n'est pas qu'une petite tache verte striée de blanc dans le coin d'un tableau et l'expression d'un visage ne se limite pas à quelques coups de crayons; il n'y a pas *identité* entre les deux même s'il y a manifestement une sorte de lien de dépendance entre les deux.

C'est ce genre de dépendance non-réductive que les philosophes cherchent à établir entre le corps et l'esprit. Comme le dit Kim: "La survenance du psychologique, si elle prévaut, nous donnerait un sens important selon lequel le physique détermine le mental: une fois que l'aspect physique de notre être est complètement déterminé, notre vie psychologique est aussi complètement déterminée. Et puisque le physique ne survient évidemment pas sur le psychologique, cette détermination est asymétrique"[10].

Trois types de survenance et leurs problèmes

J'ai déjà mentionné que la survenance se prête à de multiples interprétations; il est maintenant temps de clarifier les principaux sens attribués à la survenance et de discuter leurs implications. Kim (1984, 1987) a relevé trois principaux types de survenance dans la littérature sur le problème corps-esprit: la survenance faible, la survenance forte et la survenance globale.

La survenance faible

La survenance faible est celle endossée par Davidson. Voici la définition formelle que Kim en donne:

> **Survenance faible:** *A survient faiblement* sur *B* si et seulement si, nécessairement, pour n'importe quel *x* et *y*, si *x* et *y* possèdent les mêmes propriétés de niveau *B* alors *x* et *y*

[10] *"Psychological supervenience, if it obtains, would give us one important sense in which the physical determines the mental: once the physical side of our being is completely fixed, our psychological life is also completely fixed. Since the physical obviously does not supervene upon the psychological, this determination is asymmetric"*. Kim (1982) aussi dans Kim (1993), essai 10, p. 176.

possèdent les mêmes propriétés de niveau A - c'est-à-dire que l'indiscernabilité de niveau B implique l'indiscernabilité de niveau A.[11]

Dans cette définition, les propriétés en A sont les *propriétés survenantes* et celles en B sont les *propriétés de base*, ou les *propriétés subvenantes*. Même si la définition exprime la relation de dépendance par une biconditionnelle, il est clair que la survenance n'est pas une relation logique en elle-même, mais bien une relation ayant un statut ontologique, c'est-à-dire décrivant une relation existant entre deux propriétés ou caractéristiques du monde physique, et non pas seulement entre des concepts ou des idées.

Prenons un exemple. Disons que la célébrité survient faiblement sur le fait de pouvoir être reconnu dans la rue par au moins un million de personnes et d'être américain. La propriété d'être célèbre est la propriété survenante C, alors que la propriété de pouvoir être reconnu dans la rue par au moins un million de personnes est la propriété de base R et la propriété d'être américain est la propriété de base A. Selon notre définition, si Bill Clinton et Mickey Mouse possèdent tous deux les propriétés R et A, alors Bill Clinton et Mickey Mouse possèdent tous deux la propriété C. On ne peut pas dire que Bill Clinton est célèbre alors que Mickey Mouse ne l'est pas: l'indiscernabilité au niveau des propriétés de base ("pouvoir être reconnu dans la rue par au moins un million de personnes" et "être américain") implique l'indiscernabilité au niveau de la propriété survenante (être célèbre).

Toutefois, on peut se demander pour quelle raison la propriété C survient précisément sur les propriétés R et A. La survenance faible ne garantit ce lien entre les propriétés de base R et A et la propriété survenante C que pour *un seul monde*. Dans un autre monde possible, on peut concevoir que la célébrité survient sur le fait de pouvoir être reconnu dans la rue par au moins *deux* millions de personnes, ou d'être un personnage de dessin animé, ou d'être le Président des État-Unis, etc., ou toute combinaison de ces propriétés de base. Selon la propriété de base ou la famille de propriétés de base sur laquelle survient la propriété C pour un monde donné,

---

[11] "*A* weakly supervenes *on B if and only if necessarily for any x and y if x and y share all properties in B then x and y share all properties in A -that is, indiscernibility with respect to B entails indiscernibility with respect to A.*" Kim (1984) aussi dans Kim (1993), p. 58. Kim souligne.

Bill Clinton et Mickey Mouse seront ou ne seront pas indiscernables pour ce monde quant à leurs propriétés de base et donc partageront ou non la propriété survenante *C*. Bref, pour répondre à notre question, la propriété *C* survient faiblement sur les propriétés de base *R* et *A* parce qu'il s'agit là d'une caractéristique de ce monde particulier.

La survenance forte

Kim présente une seconde définition de la survenance faible, équivalente à la première que nous avons vue, mais qui permet de mieux voir la différence entre survenance faible et survenance forte:

> **Survenance faible (2):** *A survient faiblement* sur *B* si et seulement si, nécessairement, pour n'importe quel *x* et n'importe quelle propriété *F* de niveau *A*, si un objet *x* possède *F*, alors il existe une propriété *G* de niveau *B* telle que *x* possède *G*; et si un *y* quelconque possède *G*, il possède *F*.[12]

Par comparaison, Kim définit la survenance forte comme suit:

> **Survenance forte:** *A survient fortement* sur *B* si et seulement si, nécessairement, pour n'importe quel *x* et n'importe quelle propriété *F* de niveau *A*, si un objet *x* possède *F*, alors il existe une propriété *G* de niveau *B* telle que *x* possède *G*; et *nécessairement* si un *y* quelconque possède *G*, il possède *F*.[13]

On remarque que la seule différence entre ces deux définitions est la présence de l'opérateur modal 'nécessairement' dans la dernière partie de la définition de la survenance forte. Cet

---

[12] "*A* weakly supervenes *on B if and only if necessarily for any property F in A, if an object x has F, then there exists a property G in B such that x has G, and if any y has G it has F.*" Cette définition et la démonstration de son équivalence avec la première définition de la survenance faible se retrouve dans Kim (1984), aussi dans Kim (1993), essai 4, p. 64. Dans cette définition de la survenance faible, Kim omet de lier sa variable *x* par un quantificateur comme il le fait pourtant dans sa définition de la survenance forte. Il s'agit d'une omission que je corrige dans ma traduction. Kim souligne.

[13] "*A* strongly supervenes *on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and* necessarily *if any y has G, it has F.*" Kim (1984), aussi dans Kim (1993), essai 4, p. 65. Kim souligne.

opérateur modal supplémentaire permet à la survenance forte d'être valide pour *tous les mondes possibles* plutôt que pour un seul comme dans le cas de la survenance faible. Pour reprendre notre exemple, disons que la célébrité (propriété *C*) *survient fortement* sur les propriétés *R* et *A*. Si Bill Clinton et Mickey Mouse sont indiscernables en regard des propriétés de base *R* et *A*, ils sont *nécessairement* indiscernables en regard de la propriété *C*. C'est donc dire que la survenance de *C* sur *R* et *A* n'est plus quelque chose de propre à un seul monde comme elle l'était dans le cas de la survenance faible, mais plutôt que la relation de survenance de *C* sur *R* et *A* est valide pour tous les mondes. Évidemment, il est peu probable que quelque chose comme la célébrité survienne de façon nécessaire sur des propriétés comme *R* et *A*. Par contre, que nos états mentaux surviennent fortement sur les états physiques de notre cerveau est beaucoup plus plausible.

Puisque la survenance forte vaut pour tous les mondes alors que la survenance faible ne vaut que pour un seul monde, cela veut dire que la survenance forte implique la survenance faible, mais non l'inverse. Prenons un monde W_1 pour lequel vaut une relation de survenance faible entre une propriété de base *G* et une propriété survenante *F*. Cette relation de survenance faible pourrait ne pas valoir pour le monde W_2, par exemple. Par contre, si dans le monde W_2 une propriété *J* survient fortement sur la propriété de base *K*, cela est valable pour tous les mondes, y compris W_1.

Si les propriétés mentales sont vues comme survenant fortement sur des propriétés physiques (par exemple, les propriétés du cerveau), cela veut dire qu'une fois que les propriétés physiques d'un individu sont déterminées, sa vie psychologique l'est également. Tout comme il n'y a pas de 'travail supplémentaire' à faire pour fixer les propriétés esthétiques d'un tableau après que ses propriétés physiques soient déterminées, un hypothétique Créateur n'aurait pas de tâche additionnelle à compléter pour déterminer notre vie mentale après avoir déterminé nos caractéristiques physiques. Mais la survenance forte n'est-elle pas trop forte? Si les propriétés physiques déterminent complètement les propriétés mentales, ne peut-on pas envisager la *réduction* des propriétés mentales aux propriétés physiques? L'existence d'une telle détermination est effectivement une condition nécessaire à la réduction, mais elle n'est pas

suffisante en elle-même. Nous pourrions être incapables de formuler une théorie explicative de la survenance des propriétés mentales sur les propriétés physiques. La réduction, tout comme l'explication, sont des activités épistémologiques et elles répondent à des impératifs différents de la seule présence d'une relation de dépendance. Par contre, même sans impliquer la réduction, la survenance forte n'est pas compatible avec l'anomalisme du mental, comme l'exige par exemple le monisme anomal. C'est que la survenance forte du mental sur le physique implique l'existence de lois psychophysiques. En effet, puisque selon la définition de la survenance forte la présence de la propriété de niveau *A* garantit la présence de la propriété de niveau *B*, il est possible d'établir une loi liant ces deux propriétés. Enfin, dire que le physique détermine complètement le mental, c'est aller trop loin pour plusieurs philosophes. Ils préfèrent se rabattre sur la survenance faible, même si elle ne vaut que pour un seul monde, ou sur la survenance globale, qu'il est maintenant temps de présenter.

La survenance globale

La survenance globale a été privilégiée par certains philosophes (Horgan, Teller, Petrie) dans l'espoir d'échapper aux inconvénients liés au choix entre survenance faible et survenance forte. La survenance globale se définit comme suit:

> **Survenance globale:** "*A* survient globalement sur *B* si et seulement si deux mondes indiscernables en regard du niveau *B* sont également indiscernables en regard du niveau *A*."[14]

La différence majeure entre la survenance globale et les types de survenance que nous avons vus précédemment est que la survenance globale ne fait pas de référence explicite aux liens pouvant exister entre des propriétés particulières. Cette caractéristique de la survenance globale est d'ailleurs ce qui la rend intéressante, car une relation de dépendance entre propriétés rappelle le réductionnisme. Appliquée au problème corps-esprit, la survenance globale signifie qu'un monde physiquement identique au nôtre serait psychologiquement identique au nôtre.

---

[14] "*A globally supervenes on B just in case worlds that are indiscernible with respect to B ("B-indiscernible", for short) are also A-indiscernible.*" Kim (1984) aussi dans Kim (1993), essai 4, p. 68.

Cela revient à dire que la survenance forte implique la survenance globale: si les propriétés de niveau A sont survenantes sur les propriétés de niveau B, alors tout monde physiquement identique au nôtre sera mentalement identique au nôtre. Par contre, la survenance globale n'implique pas la survenance forte (auquel cas la survenance forte et la survenance globale serait identique). Voici la démonstration qu'en donne Kim: prenons deux mondes, W_1 et W_2, chacun d'eux comprenant deux objets ou individus, soit x et y. Dans le monde W_1, x possède les propriétés (ou groupe de propriétés) G et F alors que y possède la propriété (ou groupe de propriétés) G. Dans le monde W_2, x possède G mais non F, et y ne possède pas G. Cette situation contrevient à la survenance *forte* de la propriété F sur la propriété G car, dans le monde W_2, x possède G mais pas F. Par contre, la survenance *globale* de F sur G est toujours possible. En effet, puisque W_1 et W_2 ne sont pas identiques quant à G, ils ne peuvent pas servir à réfuter la survenance globale de F sur G. Puisque cette situation interdit la survenance forte de F sur G mais non la survenance globale de F sur G, la survenance globale n'implique pas la survenance forte.[15]

Quel genre de dépendance peut-on attendre de la survenance globale si elle n'implique pas la survenance forte? Pour reprendre l'exemple que l'on vient de voir, en quel sens F dépend-il de G compte tenu des différences entre W_1 et W_2? En effet, on a vu que la survenance globale permet que, dans le monde W_1, F survient sur G pour x, alors que dans le monde W_2 x possède G sans avoir F. Avec la survenance globale, la seule chose qui peut être invoquée pour 'expliquer' cette différence de traitement pour x selon les mondes W_1 et W_2 est que y (un objet ou individu possiblement complètement indépendant de x) ne possède pas G en W_2 alors qu'il l'avait en W_1. C'est donc dire que, pour x, la survenance de F sur G est influencée par des objets ne lui étant absolument pas reliés!

Finalement, la survenance globale n'implique pas non plus la survenance faible. Par exemple, prenons un monde W_3 où un objet ou un individu x' possède les propriétés G et F alors qu'un autre objet ou individu y' possède G mais non F. Ce monde ne permet pas la survenance faible de F sur G puisque x' et y' sont identiques quant à G mais différents quant à la propriété F. Par

[15] Cet exemple, cité par Kim (1987), aussi dans Kim (1993), essai 5, p. 82-83, est de Petrie (1987).

contre, cette situation est tout à fait compatible avec la survenance globale de F sur G, puisque la survenance globale exige seulement qu'il n'y ait pas un *autre* monde, disons W_4, où F serait distribué différemment sans que G soit distribué différemment, par exemple où x' aurait G mais non F et où y' aurait F et G. Bref, la survenance globale empêche l'existence combinée de W_3 et de W_4, mais n'empêche pas que, *dans le même monde*, deux individus identiques physiquement soient différents mentalement. Ce genre de cas intuitivement peu crédibles conduit Kim à demander: "Comment peut-on affirmer la dépendance du mental sur le physique si, comme le permet la survenance globale du mental sur le physique, il devait exister un être humain physiquement identique à vous à tous égards et ayant pourtant une vie mentale complètement différente de la vôtre, voire pas de vie mentale du tout?"[16]

La survenance globale est donc extrêmement faible. Et cette faiblesse ne semble pas permettre d'obtenir de relation de dépendance ayant une force suffisante pour être intéressante. En effet, la survenance globale peut conduire à des situations intuitivement peu crédibles, comme on vient de le voir. Une autre situation de ce genre que permet la survenance globale est qu'un monde physiquement identique au nôtre à l'exception d'un détail mineur (pour reprendre l'exemple de Kim, disons que dans ce monde les anneaux de Saturne contiennent une molécule d'ammoniac supplémentaire) diffère complètement du nôtre mentalement. La seule présence d'une molécule supplémentaire dans ce monde alternatif lui permet de manifester une absence totale de propriétés mentales, ou une distribution de ces propriétés complètement différente de celle de notre monde, comme par exemple que les créatures ayant des cerveaux ne soient pas conscientes alors que les rochers le soient.

Certes, on peut reprocher à la plupart des considérations de cette section sur la survenance globale de faire appel à des relations de survenance *de propriété à propriété* alors que la survenance globale est justement destinée à *ne pas* s'occuper de ce genre de relations particulières. Mais justement, ce que les exemples de cette section montrent, c'est que la

---

[16] "*How it is possible to advance a claim of physical dependence of the mental if, as permitted by the global supervenience of the mental on the physical, there should exist a human being physically indistinguishable from you in every respect who has a mental life entirely different from yours or who has no mental life at all?*" Kim (1987), aussi dans Kim (1993), essai 5, p. 84.

négligence des relations de survenance pouvant exister entre les propriétés particulières conduit à des situations rebutantes pour le sens commun. En fait, les cas où la survenance globale apparaît comme une bonne idée sont toujours des cas où la survenance forte vaut également et, inversement, les cas où elle apparaît absurde sont toujours des cas où la survenance forte ne vaut pas. Si les propriétés mentales surviennent globalement sur les propriétés physiques d'un monde, on s'attend naturellement à ce que, pour chaque objet, ses propriétés mentales surviennent sur ses propriétés physiques. Or, nous avons vu que la survenance globale n'implique pas la survenance forte. Et lorsque l'on examine les cas où la survenance globale tient mais où la survenance forte ne vaut pas, il devient impossible de donner une explication plausible du genre de dépendance que la survenance globale représente. En fait, on ne s'attend pas à réellement découvrir ce genre de situation dans le monde. Nous sommes toujours tentés d'expliquer un fait général ou global par l'existence des occurrences particulières de ce fait, mais cette stratégie est bloquée par la survenance globale. La survenance globale n'apparaît donc pas comme une voie prometteuse pour la résolution du problème corps-esprit. Aussi n'en sera-t-il plus question dans le reste de ce mémoire.

Perspectives sur la survenance

La seule relation de dépendance ontologique qui soit vraiment reconnue et largement discutée est la causalité. Une cause détermine son effet d'une manière qui nous apparaît naturelle (malgré les difficultés que nous éprouvons toujours à définir plus précisément cette notion). Si la survenance (forte) est introduite en tant que nouvelle forme de dépendance, il faut savoir en quoi elle nous donne une forme de dépendance qui nous apparaît valable tout en différant de la causalité.

Alors que la causalité n'inspire guère les philosophes dans leur quête d'explication des états mentaux (à témoin la tentative de Searle (1992) qui ne trouve pas beaucoup de partisans) et que la réduction est jugée trop exigeante, certains philosophes ont cru trouver dans la survenance une alternative intéressante qui permettrait de reconnaître la primauté du physique sur le mental sans conduire à la réduction. Or, une particularité de la survenance est qu'il s'agit d'une notion spécialisée qui ne se retrouve pas dans la vie courante (contrairement à la causalité). Il n'y a

pas "d'analyse du concept ordinaire" à faire, ni d'intuitions avec lesquelles s'accorder. Au-delà du sens général donné par les textes philosophiques antérieurs, chacun est libre de préciser la notion de survenance comme il l'entend, d'où l'apparition de variantes forte, faible ou globale. Il n'est donc pas étonnant qu'une telle notion rencontre la faveur des philosophes, puisqu'ils peuvent, en principe, faire correspondre la définition de la survenance à leurs besoins. Toutefois, lorsque l'on parle d'une notion de relation ontologique, c'est que l'on désire référer à une relation entre des faits (ou des événements, des propriétés, etc.) qui existent déjà dans le monde. Il s'agit d'identifier et de *décrire* cette relation, non de l'inventer. Si trop de libertés sont prises avec la définition de la relation de survenance, elle perdra en légitimité ce qu'elle gagnera en correspondance avec nos désirs. Car à quel statut ontologique peut prétendre une relation qui ne soit qu'une création *a priori* de notre esprit définie de manière à correspondre à nos convictions plutôt qu'à nos découvertes? Il ne semble pas y avoir une grande différence entre, d'une part, accepter d'emblée la compatibilité du physicalisme et du réalisme du mental sur la base de sa propre conviction de leur compatibilité et, d'autre part, créer sur mesure une relation (la survenance) qui permette de combiner ces thèses. Bref, il ne faut pas trop jouer à Humpty Dumpty avec la survenance.

Kim soutient que la relation de survenance n'est pas une relation explicative en soi; qu'elle peut bien *décrire* un état de fait, mais qu'une compréhension plus profonde nécessite le recours à d'autres types de relation (comme par exemple la causalité ou l'identité). D'ailleurs, si nous nous en tenons à la lettre des définitions de la survenance, nous apprenons que la présence d'une propriété implique la présence d'une autre propriété, mais pas la raison pour laquelle cette cooccurrence de propriétés se produit. Ainsi, en affirmant qu'une propriété A survient sur une propriété B, une relation de dépendance entre A et B est établie, mais ce qui explique cette dépendance n'est pas précisé et certaines questions restent non-répondues. Par exemple, pourquoi les propriétés mentales surviennent-elles sur les propriétés physiques? Ou même: pourquoi telle propriété mentale (par exemple, la sensation de douleur) plutôt que telle autre (disons, la sensation de picotement) survient-elle sur cette propriété physique particulière (par exemple, une activation de fibres-C)? Pour Kim, il s'agit là de questions légitimes auxquelles la relation de survenance ne peut répondre. Elle ne nous dirait pas ce qui *explique* ou *fonde* la

relation de dépendance que décrit la survenance. Si cela est vrai, on ne peut plus compter sur la survenance pour *expliquer* la compatibilité du physicalisme non-réductif et de la fermeture du monde physique. La thèse de la survenance du mental sur le physique ne serait donc pas une thèse explicative en soi. Or, une explication de ce en quoi consiste le genre de dépendance reliant le mental et le physique est justement ce que l'on attend de notre théorie du mental. Que la survenance ne soit pas une relation explicative est appuyée par le fait que plusieurs approches au problème corps-esprit acceptent (ou sont compatibles avec) la thèse de la survenance du mental sur le physique tout en étant incompatibles entre elles sur la manière exacte dont le mental dépend du physique. Pour Kim, cela signifie que l'affirmation d'une relation de survenance constitue l'affirmation du problème plutôt que sa solution: c'est cette survenance du mental sur le physique qu'il nous faut expliquer.

Pourtant, le grand intérêt que suscite la relation de survenance ne provient pas de son simple pouvoir descriptif. Plusieurs philosophes considèrent la survenance comme une relation ontologique authentiquement nouvelle et donc qui soit explicative au même titre que la causalité. S'il est acceptable de dire qu'à la question "Pourquoi A cause-t-il B?" on peut répondre "Parce que A est la cause de B", alors similairement il est possible de répondre à la question "Pourquoi la propriété *A* survient-elle sur la propriété *B*?": "Parce que *B* est la base subvenante de *A*". Ce genre de réponse n'est pas plus circulaire dans un cas que dans l'autre.

Quoiqu'il en soit, même avec l'approche de Kim, la survenance reste quelque chose de positif en ce qu'elle représente la base commune sur laquelle s'entendent toutes les approches physicalistes du mental. En effet, prise comme signifiant essentiellement que le mental est dépendant du physique, la survenance est une position qui correspond tant au monisme anomal qu'au réductionnisme et qui constitue ainsi la définition d'un "physicalisme minimal" que doivent respecter toutes les approches se disant matérialistes.

Chapitre 2: L'argument de l'exclusion

Le physicalisme non-réductif constitue présentement le champ de recherche de la grande majorité des philosophes travaillant sur la causalité mentale. Les partisans du physicalisme non-réductif se distinguent en ce qu'ils considèrent possible d'accommoder le réalisme du mental (c'est-à-dire l'attribution de pouvoirs causaux aux événements mentaux) avec le physicalisme, qui implique pourtant la fermeture du monde physique (seuls des événements physiques peuvent causer des événements physiques). J'ai déjà mentionné les raisons qui supportent l'attribution de pouvoirs causaux aux événements mentaux et celles appuyant la fermeture du monde physique. Kim est en principe d'accord avec ces deux thèses, mais il croit que leur conjonction pourrait bien être impossible.

Pour le montrer, Kim propose un argument que l'on peut nommer de façon générale "l'argument de l'exclusion". Cet argument repose essentiellement sur le fait que, d'une part, les partisans du physicalisme non-réductif permettent la formulation d'explication de nos actions sur la base de nos événements mentaux et que, d'autre part, leur acceptation du physicalisme permet également de fournir une explication des mêmes phénomènes en termes physiques. Kim suggère alors que de proposer deux explications distinctes d'un même phénomène produit une situation instable. Cette instabilité provient de ce que Kim considère que proposer deux explications différentes équivaut à proposer deux *causes* différentes d'un même phénomène. Or, s'il faut choisir entre une cause mentale et une cause physique, la cause physique l'emporte puisqu'elle ne présente pas de problème métaphysique. Sa conclusion est qu'il ne reste aux physicalistes que deux options: l'épiphénoménisme ou le réductionnisme. Cet argument de Kim se divise en deux parties. D'une part, il y a le *problème de l'exclusion causale/explicative*, destiné à montrer l'instabilité de toute double explication d'un phénomène. D'autre part, il y a ce que Kim nomme *"l'argument de la survenance"*, dont il se sert pour montrer que l'acceptation de la survenance, quoique nécessaire à tout physicaliste, conduit par ailleurs à une sévère remise en question de la valeur des explications recourant aux événements mentaux. En effet, les propriétés physiques subvenantes paraissent constituer de meilleures

candidates à la possession de pouvoirs causaux que les propriétés mentales survenantes. Je vais présenter en détail ces deux arguments complémentaires qui conduisent au dilemme de Kim et tenter d'en montrer la validité, avant de passer, au chapitre 3, aux tentatives d'explication de la causalité mentale que Kim a élaborées.

Le problème de l'exclusion causale/explicative

D'après Kim, il existe un lien entre donner une explication et identifier la cause d'un phénomène même si le premier est de nature épistémologique et le second de nature ontologique: "...pour qu'un phénomène quelconque ait un rôle explicatif, il faut que sa présence ou son absence fasse une différence - une *différence causale*."[17] Autrement dit, selon Kim, une explication (causale) valide d'un phénomène doit nécessairement référer à la *cause* de ce phénomène. Ainsi, pour que l'explication de *e* par référence à *c* constitue une explication valide, il faut que *c* soit la cause de *e*. Il s'agit là d'une vision "objectiviste" ou "réaliste" de l'explication que Kim appelle *"réalisme explicatif"* (*"explanatory realism"*). Une explication qui réfère à une cause est une *explication causale*. Mais restreindre la signification du concept d'explication à celui d'explication causale est sans aucun doute abusif. Car s'il est probable que la référence à une cause constitue une condition *suffisante* à l'obtention d'une explication, la référence à une cause n'est certainement pas *nécessaire* à toute explication. Selon les contextes, une explication satisfaisante peut consister en autre chose que l'identification d'une cause. Par exemple, si je désire expliquer ce qu'est un ordinateur, je n'ai pas besoin de référer à sa cause, c'est-à-dire à son concepteur ou à son fabricant. Expliquer *ce qu'est* quelque chose, expliquer *le but* de quelque chose ou expliquer le *pourquoi* de quelque chose constituent manifestement des significations différentes du concept d'explication. Aussi, pour ne pas se perdre dans les débats sur la nature de l'explication, Kim accepte de limiter l'application du réalisme explicatif aux seules explications causales:

---

[17] "*...for any phenomenon to have an explanatory role, its presence or absence in a given situation must make a difference - a* causal difference." Kim (1998), p. 31.

> La notion d'explication est assez vague et large [...] et personne ne devrait décider de ce qu'est une explication et de ce qui n'en est pas, à l'exception toutefois de ceci, à savoir que nous devrions insister pour que, lorsque l'on parle "d'explication causale", ce qui est désigné comme une cause doive réellement être la cause de ce qui est expliqué. Le réalisme explicatif devrait au moins s'appliquer à l'explication causale.[18]

La référence à une cause constitue une explication dans un sens important et le réalisme explicatif apparaît comme une thèse intuitive quant aux explications causales: lorsque l'on demande le *pourquoi* d'un événement, c'est que l'on veut en savoir la cause.

De plus, les explications causales jouent un rôle important dans le problème de la causalité mentale. Nous référons à nos états mentaux pour expliquer nos actions parce que nous croyons qu'ils en constituent la cause et, inversement, nous sommes portés à accepter l'attribution de pouvoirs causaux aux états mentaux parce que nous référons à eux comme explication de nos actions. La thèse du réalisme explicatif permet de rendre compte du fait que les partisans du physicalisme non-réductif considèrent les événements mentaux comme étant à la base des explications psychologiques *et* comme possédant des pouvoirs causaux. Il est vrai que les événements mentaux peuvent servir à expliquer nos actions: je lève mon bras en classe parce que je désire poser une question et que je crois qu'il s'agit là de la façon appropriée d'intervenir en classe. Mais il semble également que nos actions peuvent recevoir une explication strictement physique: je lève mon bras grâce à la contraction du muscle de mon épaule, le deltoïde, cette contraction s'expliquant elle-même par l'envoi d'un influx nerveux dans ce muscle par le biais d'un nerf reliant ce muscle à un neurone du système nerveux, la présence d'un influx nerveux dans ce neurone s'expliquant par la présence d'influx nerveux dans les neurones antérieurs, etc. Nous pouvons reculer dans les causes physiques sans jamais tomber sur un état mental. Cette explication en termes physiques de mon comportement apparaît tout à fait valable. Toutefois, nous nous trouvons alors en présence de deux explications causales différentes, l'une psychologique, l'autre physique, pour le même fait. Est-ce que cela pose un

---

[18] "*Explanation is a pretty loose and elastic notion [...] and no one should legislate what counts and what doesn't count as explanation, exepting only this, namely that when we speak of "causal explanation", we should insist, as I said, that what is invoked as a cause really be a cause of whatever it is that is being explained. Realism about explanation should at least cover causal explanation.*" Kim (1998), p. 76.

problème? Est-il acceptable de supposer que deux explications causales indépendantes puissent rendre compte d'un seul et même phénomène? Est-il généralement possible de donner plus d'une explication d'un même phénomène? Ou au contraire un événement ne reçoit-il généralement qu'une seule explication? En réponse à ces questions, Kim propose le principe suivant:

> **Principe de l'exclusion explicative:** Aucun événement ne peut recevoir plus d'une explication *complète* et *indépendante*.[19]

Compte tenu de notre discussion précédente, il faut bien sûr comprendre qu'aucun événement ne peut recevoir plus d'une explication *causale* complète et indépendante. Ce principe affirme que deux explications d'un même phénomène ne peuvent se côtoyer sans que cela crée une tension ou une instabilité entre les deux explications proposées. Si deux explications sont possibles pour expliquer le fait que je lève mon bras, ces deux explications ne peuvent coexister sans qu'une compétition explicative s'installe entre les deux. Il faut alors choisir entre les deux ou montrer en quoi ces deux explications ne sont pas complètes ou indépendantes.

Il est toutefois important de distinguer le principe de l'exclusion explicative d'un autre principe, celui de *l'exclusion causale*. Le principe de l'exclusion causale se lit comme suit:

> **Principe de l'exclusion causale:** Aucun événement ne peut avoir plus d'une cause *complète* et *indépendante*.

Le principe de l'exclusion explicative est d'ordre épistémologique alors que celui de l'exclusion causale est d'ordre ontologique. Malgré cette différence fondamentale, cependant, ces deux principes sont souvent traités par Kim comme ne constituant en fait qu'un seul et même principe, ce qui s'explique par le fait que Kim les acceptent tous les deux et qu'ils ne sont pas sans liens entre eux. En effet, il est impossible de soutenir le principe de l'exclusion

---

[19] *[The principle of explanatory exclusion] "No event can be given more than one* complete *and* independent *explanation."* Kim (1989), aussi dans Kim (1993), essai 13, p. 239.

explicative tout en rejetant le principe de l'exclusion causale. Car si un événement possèdait plus d'une cause complète et indépendante, alors il serait évidemment possible de formuler plus d'une explication (une pour chaque cause) qui soit également complète et indépendante. Qu'en est-il de l'inverse? Est-il possible d'accepter le principe de l'exclusion causale tout en rejetant celui de l'exclusion explicative? Contrairement à ce que semble croire Kim, rien ne s'y oppose de prime abord, puisque le principe de l'exclusion causale lui-même ne force aucunement l'adoption du principe de l'exclusion explicative.

D'ailleurs, alors que le principe de l'exclusion causale est largement accepté, celui de l'exclusion explicative est contesté par une majorité de philosophes. Ainsi, Tyler Burge et plusieurs autres, comme Dreske, Davidson et Van Gulik[20], refusent ce principe. Leur point central de leur position est le suivant: toute explication est fonction de nos intérêts. On peut légitimement s'intéresser aux phénomènes microscopiques ou aux phénomènes macroscopiques (dont font partie les phénomènes mentaux). Étant donné que les explications de ce qui se passe au niveau microscopique ne rendent pas compte de ce qui se passe au niveau macroscopique, les deux espèces d'explication ne sont pas rivales. Ainsi, un médecin expliquera un mouvement du corps en parlant de muscles et d'influx nerveux, alors qu'un psychologue expliquera ce comportement en parlant de cognition. Ces deux explications expliquent le même phénomène en employant des concepts différents qui relèvent de théories différentes. Malgré tout, cela ne veut pas dire que l'une de ces théories est nécessairement plus adéquate que l'autre. De plus, puisque les deux espèces d'explication ont leur utilité, nous forcer à choisir entre les deux représente une perte. En effet, il n'est pas très éclairant pour un psychologue de parler de muscles, comme il n'est pas très éclairant de parler de cognition à un médecin. Burge écrit:

> Il serait retors de penser que les explications en termes mentaux empêchent ou interfèrent avec les explications non intentionnelles du mouvement physique. [...] [Ces idées] semblent retorses parce que nous savons que les deux explications causales expliquent le même effet physique en passant par deux configurations d'événements très différentes. Les explications de ces configurations répondent à deux espèces de questionnement très

[20] Dreske (1981), p. 211, Davidson (1993), p. 16 et Van Gulik (1993), p. 255-256.

différentes. Aucune des deux espèces d'explication ne fait de suppositions essentielles ou précises à propos de l'autre.[21]

Burge parle bien ici d'explication causale, ce n'est donc pas là l'origine de sa mésentente avec Kim. De plus, rien ne semble s'opposer à ce que Burge et les autres philosophes qui rejettent le principe de l'exclusion explicative acceptent par ailleurs le réalisme explicatif (la thèse selon laquelle une explication causale valide d'un phénomène doit nécessairement référer à la cause de ce phénomène). Aussi, lorsque Burge déclare qu'il peut y avoir plus d'une explication d'un même phénomène, il ne veut pas dire que ces explications réfèrent à des causes différentes. En effet, si on accepte le principe de l'exclusion causale et le réalisme explicatif, alors deux explications d'un même phénomène doivent nécessairement référer à la même cause. Pour pouvoir rejeter le principe de l'exclusion explicative dans ces conditions, il faut donc que deux explications causales référant à la même cause puissent être différentes. Pour plusieurs philosophes comme Burge, ce qui différencie une explication d'une autre ce n'est pas son référent, mais sa signification. Deux explications d'un même phénomène ont les mêmes référents: elles réfèrent aux mêmes événements (le même effet et la même cause). Mais elles n'ont pas la même signification: elles ne font pas partie de la même théorie, n'utilisent pas le même vocabulaire, ni les mêmes concepts. Il est facile de voir que l'on peut *référer* à la même chose sans *signifier* la même chose: Si je dis que Oedipe a épousé Jocaste, ou que Oedipe a épousé sa mère, je réfère à la même chose, puisque Jocaste est la mère de Oedipe. Mais je ne signifie pas la même chose, puisque l'on peut ne pas savoir que Jocaste est la mère d'Oedipe. D'ailleurs, si Oedipe avait su que Jocaste était sa mère, il ne l'aurait pas épousée. Que deux explications soient différentes n'excluent pas la possibilité qu'elles puissent se ramener l'une à l'autre par une réduction. Mais *même si ce n'est pas le cas* (comme, selon eux, dans le cas des états mentaux), Burge et les autres philosophes partageant son avis (appelons-les les "anti-

---

[21] "*It would be perverse to think that the mentalistic explanation excludes or interferes with non-intentional explanation of the physical movement. [...] They seem perverse because we know that the two causal explanations are explaining the same physical effects as the outcome of two very different patterns of events. The explanations of these patterns answer two very different types of inquiry. Neither type of explanation makes essential, specific assumptions about the other*" Burge *in* Heil et Mele (1993), p. 116, cité par Kim (1998), p. 65.

exclusionnistes") considèrent que ces explications peuvent coexister sans difficulté, qu'elles ne sont aucunement rivales.

Kim est bien sûr de l'avis contraire. Il soutient qu'on ne peut séparer aussi facilement les considérations épistémologiques des considérations métaphysiques. Par exemple, la simple affirmation que les états mentaux (croyances, désirs, etc.) peuvent servir à formuler des explications *causales* est une affirmation métaphysique. Or, s'il est impossible de séparer les considérations épistémologiques des considérations métaphysiques, on ne peut pas non plus séparer le principe de l'exclusion explicative du principe de l'exclusion causale. Pour cette raison, Kim considère que deux explications différentes d'un même phénomène réfèrent nécessairement à deux causes différentes, et qu'en vertu du principe de l'exclusion causale, elles sont en compétition l'une avec l'autre. Bref, pour Kim, la valeur du principe de l'exclusion explicative découle de celle du principe de l'exclusion causale et ce, parce les considérations épistémologiques ne peuvent être séparées des considérations métaphysiques comme le font les anti-exclusionnistes qui ne voient pas en quoi deux explications d'un même phénomène peuvent être rivales si elles répondent à des questions différentes et qu'elles relèvent de théories différentes. C'est donc en ayant en tête tant des considérations épistémologiques que métaphysiques que Kim écrit en réponse à Burge dans *Mind in a Physical World* (1998):

> Ce que Burge rate d'intéressant à propos des explications, c'est que *deux explications ou plus peuvent être des explications rivales même si leurs prémisses explicatives sont mutuellement consistantes et tout à fait véridiques, dès lors qu'elles visent à expliquer (plus précisément, causalement expliquer) un explanandum unique*. Que ces explications se présentent dans des domaines de recherche différents, qu'elles soient données à des 'niveaux' d'analyse ou de descriptions différents ou qu'elles répondent à des considérations épistémologiques ou pragmatiques différentes ne fait pas de différence.[22]

---

[22] *"The interesting fact about explanations that Burge misses is that* two or more explanations can be rival explanations even though their explanatory premises are mutually consistent and in fact all true, if they purport to explain (in particular, causally explain) a single explanandum. *That the explanations arise in different areas of inquiry, that they are given at "different levels" of analysis or description, or that they are responses to different epistemic or pragmatic concerns, makes no difference.*" Kim (1998), p. 65-66. Kim souligne.

Qui a raison alors? Deux explications causales d'un même phénomène sont-elles ou non rivales? Je ne prétends certainement pas être en mesure de règler ce litige concernant la nature des explications causales. Néanmoins, j'aimerais faire remarquer qu'il fait assurément sens de dire que deux explications n'ayant pas la même signification sont différentes. De même qu'il est compréhensible de dire que deux théories ne sont pas nécessairement réductibles l'une à l'autre même si elles réfèrent aux mêmes phénomènes; leurs concepts peuvent être si différents qu'il soit impossible de ramener ceux d'une théorie à ceux de l'autre. Il ne faut donc pas se surprendre si deux explications relevant de deux théories différentes ne sont pas réductibles l'une à l'autre. Par contre, il est clair que la vertu explicative d'une explication causale relève de manière importante de la référence à une cause et n'est donc pas exclusivement tributaire de sa signification. Le principe de l'exclusion explicative de Kim a le mérite de rendre compte de l'intuition selon laquelle deux explications ne nous apparaissent pas fondamentalement différentes si, au bout du compte, elles sont réputées référer à la même cause.

Quoiqu'il en soit, il me semble qu'il pourrait bien s'agir d'un faux débat, du moins quant à l'opposition existant entre Kim et ceux qui considèrent que plusieurs explications d'un même événement ne sont pas rivales, comme Burge, Dreske, Davidson et Van Gulik. J'entends montrer que malgré les désaccords théoriques existant entre les deux camps, Kim reconnaît en fait la possibilité que défendent les anti-exclusionnistes. Pour défendre cette affirmation de ma part, il est nécessaire de passer en revue les différents cas de figures que voit Kim lorsque deux explications différentes sont proposées pour rendre compte d'un même événement. La particularité de cette présentation de Kim, c'est qu'il parle de deux explications *référant à des causes différentes*. Or, on l'a vu, pour les anti-exclusionnistes, deux explications différentes rendant compte d'un même phénomène réfèrent *à la même cause* (autrement, ils ne respecteraient pas le principe de l'exclusion causale). Pour voir si Kim et ces philosophes s'opposent réellement, il nous faut donc examiner les cas où les deux explications réfèrent à la même cause.

Cas de figures face à l'exclusion explicative

Supposons alors qu'une explication *C* réfère à la cause *c* pour expliquer un événement *e* donné et qu'une explication *C** réfère à la cause *c** pour expliquer ce même événement. On retrouve chez Kim les cinq possibilités principales[23] suivantes:

Premier cas. Ni *c*, ni *c** n'est en elle-même une cause suffisante de *e*, mais chacune d'elles constitue un composant nécessaire à l'apparition de *e*. L'apparente indépendance de *c* et *c** provient de ce que nous ne mentionnons souvent que l'élément, *c* ou *c** qui nous apparaît le plus pertinent dans les circonstances, sur la base de considérations épistémologiques ou pragmatiques. Par exemple, un accident d'automobile peut s'expliquer par la chaussée glissante, la mauvaise condition des freins et l'inexpérience du conducteur. Aucun de ces facteurs n'aurait pu causer l'accident à lui seul, mais leur combinaison le peut. Pourtant, selon le contexte, nous pouvons ne mentionner que l'un ou l'autre de ces facteurs en guise d'explication. Toutefois, *c* et *c** ne constituent pas alors deux causes *complètes* de *e*, mais bien deux causes *partielles*. Et puisque l'explication *C* et l'explication *C** ne réfèrent qu'à des causes partielles, elles ne constituent elles-mêmes que des explications partielles et peuvent ainsi coexister sans créer de tension.

Un autre cas d'explication partielle est celui où, plutôt que *c* et *c** soient chacun une cause partielle de *e*, *c* se révèle être une partie de *c**. La cause *c**, puisqu'elle comprend *c*, est donc bien une cause complète de *e*. Mais *c* elle-même ne constitue qu'une cause partielle de *e* et ce, même s'il peut être pertinent dans certains contextes de mentionner *c* comme cause de *e*. Une fois que *c* est reconnue comme n'étant qu'une simple partie de *c**, la tension est résolue entre les explications *C* et *C**.

Second cas. Les causes *c* et *c** sont deux maillons différents d'une même chaîne causale menant à *e*. Disons que la chaîne causale est la suivante: *c* cause *c**, laquelle cause ensuite *e*. Puisque *c* conduit à la présence de *c**, laquelle cause *e*, *c* peut être considérée comme la cause *indirecte* de *e*, ce qui permet à l'énoncé "si *c* ne se produisait pas, *e* ne se produirait pas" d'être

[23] Voir Kim (1993), essai 13, p. 250-252 et Kim (1998), p. 64-65.

vrai. La causalité est fréquemment définie par de tels énoncés contraires aux faits (*'contrefactual'*), ce qui rend compréhensible que quelqu'un fasse référence à *c* comme étant la cause de *e*. Malgré tout, il est faux de dire que *c* constitue une cause de *e* qui soit complète et indépendante, puisqu'elle nécessite la présence de l'événement *c** comme intermédiaire pour pouvoir expliquer *e* de façon appropriée. Prenons un exemple. Un joueur de billard frappe sa boule blanche (événement *c*) qui entre ensuite en collision avec une autre boule (événement *c**), laquelle tombe dans une poche (événement *e*). Il est plausible qu'un spectateur explique l'événement *e* (faire tomber la seconde boule dans la poche) en mentionnant simplement l'explication *C*, c'est-à-dire en référant exclusivement à la cause *c* (le fait de frapper la boule blanche) et en négligeant l'explication *C** qui réfère à la cause intermédiaire *c** (la boule blanche frappant la seconde boule). On peut même ajouter un certain nombre d'étapes intermédiaires (d'autres boules impliquées et quelques rebonds sur les bords) pour amplifier l'aspect *partiel* de l'événement *c* (et des autres étapes éventuellement impliquées) quant à la production de l'événement *e* sans changer fondamentalement le problème. Reste que sans le dernier événement *c**, *c* ne peut conduire à *e*. La cause *c* n'est donc pas une cause complète et indépendante de *e*.

Les deux prochains cas sont particulièrement intéressants lorsque nous les plaçons en perspective avec ce que disent les anti-exclusionnistes.

Troisième cas. Nous trouvons que les causes *c* et *c** sont identiques *et* que les deux explications se ramènent à l'une à l'autre. L'apparence de dualité provient simplement de ce que les explications *C* et *C** correspondent à des descriptions différentes de la même cause. L'une des explications correspond alors à une description plus profonde, plus systématique et plus abstraite que l'autre de la même cause. Tel est le cas de l'exemple de la température donné précédemment pour expliquer le réductionnisme: décrire la température en termes d'énergie cinétique des molécules d'un corps offre une vision plus profonde (phénomène microscopique), plus systématique (recours à des lois statistiques) et plus abstraite (concept d'énergie cinétique) que de simplement parler de la température d'un corps (simple échelle graduée). Malgré tout, la température et l'énergie cinétique des molécules d'un corps n'en sont

pas moins identiques. Expliquer la liquéfaction d'un cube de glace par une augmentation de température de l'eau ou une augmentation de l'énergie cinétique des molécules d'H_2O, c'est référer à la même cause. Ainsi, lorsqu'il y a identité de *c* et de *c**, nous n'avons pas deux explications causales de *e* mais une seule. Pour Kim, il n'y a alors qu'une seule explication et le principe de l'exclusion explicative est respecté.

Ce cas de figure me semble correspondre à ce dont parlent Burge et les autres anti-exclusionnistes. En effet, ceux-ci considèrent que les deux explications réfèrent à la même cause? C'est ce que fait Kim ici. Ces philosophes ne voient pas de rivalité entre de telles explications rendant compte d'un même phénomène? Kim non plus. La seule différence entre les deux camps est que les anti-exclusionnistes maintiennent que les deux explications sont distinctes, alors que Kim les assimile. Et ce curieux accord ne dépend nullement du fait que l'explication *C* se réduise à l'explication *C**. Les mêmes commentaires que je viens de faire s'appliquent tout aussi bien si *C* et *C** ne sont pas réductibles l'un à l'autre, à condition que *C* et *C** réfèrent à la même cause. C'est d'ailleurs là le prochain cas que distingue Kim.

Quatrième cas. Nous trouvons que les causes *c* et *c** sont identiques, mais que malgré tout l'explication *C* et l'explication *C** ne sont pas réductibles l'une à l'autre. Il peut s'agir de cas similaires à ceux que l'on vient de discuter mais où, pour quelque raison que ce soit, la réduction est impossible. La survenance psychophysique est souvent perçue comme un tel cas, puisqu'elle s'oppose à la réduction. Quoiqu'il en soit, nous n'avons pas affaire à deux causes indépendantes de *e*.

Il s'agit bien ici de la situation privilégiée par les anti-exclusionnistes: deux explications référant à la même cause, mais qui ne sont pas réductibles l'une à l'autre. Pour Kim, cela signifie que les explications *C* et *C** ne sont pas indépendantes. Pour les anti-exclusionnistes, au contraire, il s'agit bien de deux explications différentes. Tant Kim que les anti-exclusionnistes, cependant, s'entendent sur une chose: il n'y a pas de rivalité entre les deux explications. Le débat entre Kim et les anti-exclusionnistes apparaît ainsi basé sur un malentendu. Kim s'oppose à ce que soient fournies plusieurs explications d'un même événement, car pour lui cela signifie que plusieurs *causes* sont réputées produire un même

événement. Mais proposer plusieurs causes, c'est contrevenir au principe de l'exclusion causale, pas au principe de l'exclusion explicative. Et puisque les anti-exclusionnistes acceptent tout comme Kim le principe de l'exclusion causale (ce que ne semble pas réaliser Kim), le seul point en litige m'apparaît concerner la définition de l'explication causale. Lorsque Kim insiste pour qu'une explication causale réfère bien à la cause de l'événement expliqué, il s'attaque à un point déjà gagné auprès des anti-exclusionnistes. Ce que Kim doit faire, s'il veut imposer son principe de l'exclusion explicative, c'est montrer que deux explications référant à la même cause sont assimilables l'une à l'autre, ou du moins qu'elles ne sont pas indépendantes. Or, jusqu'à maintenant, aucun argument n'a été apporté par Kim sur ce point, à moins que l'on puisse considérer comme allant dans ce sens son appel à la simplicité: "il est raisonnable de penser que l'existence de plusieurs explications pour un *explanandum* unique est fort probablement contre-productif en vue d'un objectif de simplification et d'unification."[24] Reste qu'il serait sans doute plus facile pour Kim d'abandonner son principe de l'exclusion explicative pour se concentrer sur celui de l'exclusion causale. La défense en serait beaucoup plus aisée et c'est tout ce dont Kim a besoin pour que porte son 'argument de la survenance', que nous verrons dans la seconde partie de ce chapitre. Avant d'y être, cependant, il reste un dernier cas de figure que distingue Kim lorsque plus d'une explication est proposée pour expliquer un phénomène.

Cinquième cas. Les explications C et C^* réfèrent chacune à une cause indépendante et complète de e, c'est-à-dire que c et c^* constituent chacune une cause suffisante de e. Autrement dit, c et c^* font partie de chaînes causales différentes mais qui aboutissent au même résultat, e. Il s'agit donc d'un cas de *surdétermination causale*: même si c n'était pas présent ou ne causait pas e, c^* causerait e et inversement, si c^* n'était pas là, c causerait e. L'exemple classique de surdétermination est celui de l'individu tué par *deux* balles au coeur parfaitement synchronisées. Quelle est alors la cause de la mort? On ne peut pas vraiment dire que la mort a été causée par la présence de *deux* balles puisqu'une seule aurait suffi à causer la mort. D'un

---

[24] "*It makes sense to think that multiple explanations of a single explanandum are presumptively counterproductive in regard to the goal of simplification and unification.*" Kim (1989a), aussi dans Kim (1993), essai 13, p. 254.

autre côté, on ne peut pas non plus identifier *laquelle des deux balles* est dans ce cas responsable de la mort, puisqu'elles sont simultanées. Il existe donc bien un cas où tant le principe d'exclusion explicative que le principe d'exclusion causale ne s'appliquent pas. On ne peut pas ramener *c* et *c** à une seule cause dans un cas de surdétermination causale, et donc la tension entre les explications *C* et *C** ne peut être résorbée.

Dans "Mechanism, Purpose, and Explanatory Exclusion"[25], Kim propose que *C* et *C** ne représentent peut-être pas deux explications complètes et indépendantes. En effet, une explication de *e* qui ne mentionnerait que *c* ou *c** ne donnerait qu'une vision partielle des faits. Cette proposition présente une certaine plausibilité (un médecin légiste qui ferait l'autopsie de notre homme tué par deux balles mentionnerait certainement les deux balles comme constituant la cause de la mort), mais correspond à un sens différent de "complet" que celui employé jusqu'ici. Alors que dans tous les cas précédents une explication était considérée comme *complète* si elle référait à une cause *suffisante* de *e*, ce n'est manifestement pas le cas ici puisque *c* et *c** sont réputées suffisantes pour *e ex hypothesis*. Le terme 'complet' semble donc avoir une signification plus large que celui de 'suffisant'. On pourrait plausiblement soutenir, me semble-t-il, qu'une cause complète d'un événement en est nécessairement une cause suffisante, mais qu'une cause suffisante d'un événement n'en est pas nécessairement une cause complète, comme le démontrerait l'exemple de la surdétermination.

Quoiqu'il en soit, Kim dispose d'un autre argument pour écarter la surdétermination comme possibilité viable dans le cas particulier des états mentaux. C'est que la surdétermination a toujours été perçue comme un cas *d'exception* (ainsi on ne trouve pas très fréquemment de mort par deux balles simultanées au coeur). Règle générale, les événements ne sont le résultat que d'*une seule cause* immédiate. Ainsi, supposer que *toute occurrence de causalité mentale* est un cas de surdétermination apparaît franchement improbable: ce serait rendre toute une classe d'événements tributaire d'un cas d'exception. Étant donné la rareté des cas de surdétermination, le principe d'exclusion causale n'est pas considéré comme étant mis en péril par cette possibilité.

---

[25] Kim (1989a), aussi dans Kim (1993), essai 13, p. 252.

Ces cinq cas de figures ne nous ont pas seulement été utiles pour clarifier les positions respectives de Kim et des anti-exclusionnistes quant au principe d'exclusion explicative, ils nous donnent également un panorama des possibilités qui s'offrent à nous en vertu du principe d'exclusion causale lorsque plus d'une cause est réputée produire un même événement. Pour obtenir le dilemme de Kim, cependant, il faut également voir comment le principe de l'exclusion causale nous contraint à choisir entre le réductionnisme et l'épiphénoménisme. C'est là qu'intervient l'argument de la survenance.

L'argument de la survenance

L'argument de la survenance se base essentiellement sur le principe de l'exclusion causale. Ce dernier nous dit que la présence de deux causes concurrentes pour le même événement crée une tension. D'un autre côté, toutes les approches soutenant le physicalisme non-réductif soutiennent également la survenance du mental sur le physique. Le fonctionnalisme, il est vrai, soutient plutôt ce que l'on pourrait qualifier de 'réalisationnisme', c'est-à-dire la thèse selon laquelle les états mentaux sont réalisés physiquement de façon multiple. Mais cela revient au même: puisque pour les fonctionnalistes les propriétés mentales sont réalisées par les structures physiques particulières d'un individu, il en découle que si deux individus sont physiquement indiscernables, alors ils seront aussi psychologiquement indiscernables car les mêmes entrées donneront lieu aux mêmes sorties. Bref, le psychologique *survient* sur le physique. Ainsi, en vertu de la survenance psychophysique, chaque état mental dispose d'une base ou d'un fondement physique, et ce pour toutes les approches du physicalisme non-réductif. Ce fait est considéré comme ce qui rend ces approches conformes au physicalisme. Par conséquent, pour toute cause mentale expliquant un événement donné, il existe une base subvenante à cette cause mentale. De plus, la thèse de la fermeture causale du monde physique nous dit que seuls les événements physiques peuvent causer des événements physiques. La question suivante se pose alors: si un événement x peut causer un événement y en vertu de ses propriétés physiques, x peut-il également causer y en vertu de ses propriétés mentales survenantes? L'argument de la survenance, comme on va le voir, répond négativement à cette question. En résumé, il a pour conséquence que, si un événement mental et son réalisateur physique sont en compétition pour

rendre compte d'un phénomène (qu'il s'agisse d'un phénomène physique ou mental), l'événement physique subvenant se révèle une cause plus plausible que l'événement mental. Cette conclusion, cependant, conduit directement à l'épiphénoménisme, la thèse selon laquelle les événements mentaux n'ont aucun pouvoir causal et à laquelle s'oppose le physicalisme non-réductif.

Dans plusieurs des écrits antérieurs de Kim, on retrouve une version abrégée de l'argument de la survenance dans le cadre de la discussion du problème de l'exclusion. Dans *Mind in a Physical World*, toutefois, Kim présente l'argument de la survenance de façon distincte sous la forme d'un argument par l'absurde visant à démontrer qu'il est tout aussi impossible de comprendre l'idée de la causalité mentale si on accepte la survenance psychophysique que si on la rejette. Je vais donc exposer cet argument en reprenant les dix étapes qu'il comprend dans la présentation de Kim[26] et je commenterai chacune de ces étapes, reprenant fréquemment les commentaires de ce dernier, tout en ajoutant certaines remarques lorsque cela m'apparaît nécessaire.

La première étape consiste bien sûr à poser l'alternative suivante:

> (i) La survenance psychophysique vaut ou elle ne vaut pas.

Si je reformule la thèse de la survenance pour l'appliquer au mental, nous obtenons:

> **Survenance psychophysique:** Les propriétés mentales surviennent sur les propriétés physiques si et seulement si, nécessairement, pour n'importe quel individu et n'importe quelle propriété mentale *M*, si un individu possède *M*, alors il existe une propriété physique *P* telle que cet individu possède *P*; et (nécessairement - s'il s'agit de survenance forte) si un quelconque individu possède *P*, il possède *M*.

Cette formulation est absolument parallèle à celle utilisée pour comparer la survenance faible et la survenance forte dans le chapitre 1. L'argument de la survenance vaut ainsi tant pour la survenance faible que pour la survenance forte. Ce que cette première étape (i) nous dit, c'est

---

[26] Kim (1998), p. 38-46.

que nous avons deux possibilités: soit les propriétés mentales disposent d'une base physique, soit elles n'en disposent pas. Voyons d'abord la seconde possibilité:

> (ii) Si la survenance psychophysique ne vaut pas, il n'y a apparemment aucune façon de rendre compte de la possibilité de la causalité mentale.

Cette affirmation découle de l'acceptation par les physicalistes de la fermeture causale du monde physique selon laquelle seuls les événements physiques peuvent causer d'autres événements physiques. En effet, sans la survenance pour garantir l'ancrage physique des propriétés mentales, celles-ci n'auraient plus rien de physique. Toute interaction entre des événements mentaux et des événements physiques contreviendrait alors à la thèse de la fermeture causale du monde physique. Puisqu'aujourd'hui la grande majorité des philosophes sont physicalistes, l'idée selon laquelle il est impossible de rendre compte de la causalité mentale sans la survenance est largement acceptée. Reste donc à voir si l'acceptation de la survenance, pour sa part, conduit également à des difficultés. Commençons donc par l'hypothèse que la causalité mentale est possible:

> (iii) Supposons que l'exemplification d'une propriété mentale *M* cause l'exemplification d'une autre propriété mentale *M**.

Cependant, puisque nous acceptons la survenance psychophysique, l'énoncé suivant, qui en est une application particulière, vaut également:

> (iv) Cette exemplification de *M** possède une base physique subvenante *P**.

Maintenant, compte tenu de (iii) et de (iv), une question importante se pose: d'où provient exactement cette exemplification de *M**? Il semble qu'il y ait deux réponses possibles à cette question:

> (v) *M** est présent à ce moment: (a) parce que selon notre hypothèse (iii), *M* a causé une exemplification de *M**; ou (b) parce que *P**, la base subvenante de *M**, est présente à ce moment.

Nous nous retrouvons donc avec deux réponses différentes et incompatibles explicant la présence de *M**. La force de ces deux réponses n'est pas la même, cependant, puisque la thèse de la survenance fait pencher la balance en faveur de la base physique de *M**, c'est-à-dire *P** dans ce cas-ci. En effet, la survenance garantit que la présence de *P** implique la présence de *M**. *P** est donc suffisant pour *M**. De plus, s'il est vrai que *P** lui-même n'est pas nécessaire pour *M** (*M** pourrait survenir sur une autre base que *P**), il reste que sans la présence de *P** ou d'une autre base physique sur laquelle surviendrait *M**, il est impossible d'avoir *M**. Par comparaison, et pour les mêmes raisons, *M* n'est ni suffisant, ni nécessaire pour *M**. Il devient donc difficile d'admettre (iii), c'est-à-dire que *M* a causé *M**. Toutefois, on pourrait être tenté de sauver *M* comme cause de *M** en affirmant que:

(vi) *M* a causé *M** *en causant P**.

(vi) expliquerait comment *M* a produit cette exemplification de *M**. Selon Kim, (vi) serait une illustration d'un principe plus général qui justifierait (vi) *même pour qui ne voit pas de tension dans* (v). Voici ce principe que j'appellerai "principe de la primauté de la base": "*Pour causer l'instantiation d'une propriété survenante, il faut causer l'instantiation de sa (ou d'une de ses) propriété(s) de base.*"[27] Ainsi, pour soulager notre mal de tête, nous prenons une aspirine: c'est-à-dire que nous agissons causalement sur les processus de notre cerveau sur lesquels notre mal de tête survient. Autre exemple: pour rendre une peinture plus belle ou plus expressive, il faut faire un travail physique sur la toile et ainsi affecter la base subvenante des propriétés esthétiques que nous voulons changer. Comme l'affirme Kim, on ne peut pas agir directement sur les propriétés esthétiques; il faut changer la toile physiquement si on veut la changer esthétiquement. Il n'y a aucun autre moyen.

Du moins, c'est là ce qu'affirme Kim. Mais imaginons qu'on me donne un placebo plutôt qu'une aspirine et que mon mal de tête disparaisse, que doit-on alors conclure? En supposant que le seul passage du temps n'est pas responsable de la disparition de mon mal de tête (c'est-

[27] "To cause a supervenient property to be instantiated, you must cause its base property (or one of its base properties) to be instantiated." Kim (1998), p. 42. Kim souligne.

à-dire que mon mal de tête serait encore là si je n'avais pas pris le placebo), il faut donc attribuer la disparition du mal de tête à l'effet placebo lui-même, autrement dit à la croyance que j'ai pris une aspirine. On pourrait donc soutenir que c'est un état mental qui est responsable de la disparition de ma douleur, et non une quelconque action sur le réalisateur physique de ma douleur. Dans le même ordre d'idée, plutôt que de retoucher ma toile, je pourrais partir du principe que "*beauty is in the eye of the beholder*" (la beauté est dans l'oeil de l'observateur) et m'arranger pour changer la perception que l'observateur a de ma toile (soit en lui expliquant la signification de ma toile, soit en l'exposant fréquemment à mon tableau jusqu'à ce qu'il en prenne l'habitude, etc.). Là encore, il n'y a pas d'action sur la base subvenante des propriétés esthétiques de ma toile (que cette base se limite à la toile elle-même ou qu'elle inclut le cerveau de l'observateur: je ne lui ai pas administré de drogues hallucinogènes!). Ce que ces contre-exemples démontrent, c'est que la primauté de la base subvenante illustrée en (vi) *n'est pas évidente en soi.* Bref, contrairement à ce qu'affirme Kim, il faut déjà être sensible à la tension présente dans (v) pour que le principe de la primauté de la base soit convaincant.

Par contre, ces contre-exemples *ne désamorcent pas la tension dans* (v), *ni ne réfutent l'avantage explicatif que présente la base subvenante par rapport à la propriété survenante.* En effet, il est toujours possible de soutenir l'alternative (b) de (v), et par conséquent de supposer (vi). Ainsi, on peut soutenir que si la croyance d'avoir pris une aspirine a fait disparaître mon mal de tête, c'est que cette croyance agit sur la base subvenante de mon mal de tête et non pas directement sur lui. De même pour le tableau: on peut soutenir que les propriétés esthétiques d'un tableau surviennent sur les caractéristiques physiques du tableau *et* sur les préférences de l'observateur; en agissant sur ces dernières, j'ai changé la base subvenante des propriétés esthétiques du tableau. Or, étant donné la tension présente dans (v) et le grand avantage explicatif que possède la base subvenante P^* par rapport à M pour rendre compte de M^* (la présence de P^* est suffisante pour M^* et la présence d'une base subvenante est nécessaire pour M^*), il semble que l'on puisse conclure à la véracité du principe de la primauté de la base. Enfin, le principe de la primauté de la base, à son tour, entraîne cette conclusion à laquelle Kim arrive également: "*sous l'hypothèse de la survenance*

psychophysique, la causalité du mental au mental implique ou présuppose la causalité du mental au physique."[28]

Continuons maintenant avec (vi). Étant donné la thèse de la survenance psychophysique, nous pouvons dire à propos de *M* que:

(vii) *M* lui-même possède une base subvenante *P*.

Le problème de l'exclusion se présente alors de nouveau, mais cette fois par rapport à *P**. D'après (vi), *M* cause *P**. Il y a toutefois de bonnes raisons de considérer que *P* cause *P**. Pour commencer, puisque *P** ne serait pas présent sans la présence de *M* et que *M* a besoin de la présence de *P* pour être là, on peut dire que *P** ne pourrait pas être là sans *P*.

De plus, on ne peut pas régler le problème en disant que *P* et *M* font partie de la même chaîne causale, c'est-à-dire que *P* causerait *M*, qui ensuite causerait *P**. En effet, la relation entre les propriétés subvenantes et survenantes d'un objet ne peut se comprendre aisément comme étant une relation de causalité. D'une part, la relation de causalité implique généralement que la cause précède l'effet, alors que dans le cas de la survenance, les propriétés subvenantes et survenantes sont simultanées. D'autre part, on peut généralement ajouter des maillons à une chaîne causale. Mais quel genre de maillons pourrait-on placer entre une propriété survenante et sa base? Aucun, semble-t-il. Finalement, le scénario de la chaîne causale implique que *M* cause *P** indépendamment de sa base subvenante *P* (comme le suppose également (vi)), mais cette possibilité n'est pas permise par la fermeture causale du monde physique qu'acceptent les physicalistes.

On ne peut pas non plus régler le problème de l'exclusion entre *P* et *M* en supposant qu'ils constituent chacun une *cause partielle* de *P** et qu'ensemble ils constituent une cause suffisante de *P**. Pour que l'hypothèse des causes partielles puisse fonctionner, il faudrait que *M* contribue de manière nécessaire à l'apparition de *P**. Cela implique que, pour

---

[28] "Under the mind-body supervenience assumption, mental-to-mental causation implies, or presupposes, mental-to-physical causation." Kim (1998), p. 43. Kim souligne.

complètement expliquer la présence de P^*, il faudrait mentionner un agent non-physique, M, en plus de P, ce qui contrevient à la fermeture causale de monde physique. De plus, même si on acceptait de ne pas tenir compte de la fermeture du monde physique, on ne pourrait pas considérer M et P comme des causes partielles. En effet, d'après (iii) M serait une cause suffisante à la présence de P^* et, d'après le principe de la fermeture du monde physique, P serait également une cause suffisante de P^*. On ne peut voir alors en quoi l'union de M et de P serait plus efficace causalement que M seul ou P seul.

Enfin, on ne peut non plus considérer qu'il s'agit d'un cas de surdétermination, c'est-à-dire que P^* est surdéterminé par deux causes suffisantes, soit M et P. Outre la difficulté déjà mentionnée, à savoir que tous les cas de causalité mentale deviennent des cas de surdétermination, cette hypothèse se bute au fait que d'avoir une cause physique disponible à chaque fois rend inutile la reconnaissance d'une cause mentale, et que d'envisager l'existence d'une cause complètement mentale contrevient à la fermeture du monde physique.

Si l'on désire rester fidèle à l'idéal non-réductionniste, la seule option restante semble alors être celle-ci:

> (viii) P cause P^*, tandis que M survient sur P et que M^* survient sur P^*.

Par conséquent, (iii), (v-(a)) et (vi) sont faux et il faut conclure que:

> (ix) La relation causale entre M et M^*, de même que celle entre M et P^* ne sont qu'illusoires et dérivent de la véritable relation causale existant entre P et P^*.

Cependant, si les relations causales entre M et M^* de même qu'entre M et P^* ne sont qu'illusoires, alors nous n'arrivons pas à rendre compte des cas de causalité mentale. En effet, M et M^* ne sont alors que des épiphénomènes, n'ayant aucun pouvoir causal de quelque sorte que ce soit, ne pouvant agir ni sur des événements physiques, ni sur d'autres événements mentaux (car nous avons vu qu'en vertu de la thèse de la survenance - nécessaire à tout physicaliste - l'action sur des événements mentaux présuppose l'action sur des événements physiques). D'où la conclusion de cet argument de la survenance:

(x) Si la survenance psychophysique ne vaut pas, la causalité mentale est inintelligible; si elle vaut, elle est également inintelligible. Par conséquent, la causalité mentale est inintelligible.

Les conséquences de l'argument

L'argument de la survenance est très important car il s'attaque aux racines mêmes du physicalisme non-réductif. Il montre l'incompatibilité de deux thèses constitutives du consensus du physicalisme non-réductif, à savoir le physicalisme et le réalisme du mental. Ainsi, à partir de la thèse du physicalisme qui force l'adoption de la survenance du mental sur le physique, nous en arrivons à conclure à l'épiphénoménisme qui, lui, s'oppose absolument au réalisme du mental en n'accordant aucun pouvoir causal aux états mentaux. La conclusion à laquelle nous conduit l'argument de la survenance n'est pas négligeable: *le physicalisme non-réductif n'est pas une position défendable*. Une conclusion aussi grave, surtout compte tenu du fait que le physicalisme non-réductif constitue la position dominante actuellement dans le débat sur la causalité mentale, exige un examen minutieux de l'argument.

On pourrait objecter à l'argument de la survenance qu'un argument similaire mènerait à refuser l'attribution de pouvoirs causaux à toutes les propriétés survenantes quelles qu'elles soient, comme, par exemple, les propriétés biologiques ou géologiques qui surviennent sur les propriétés microphysiques. Une telle conclusion nous apparaît absurde et nous conduit à douter de la valeur de l'argument de la survenance. Bien que la dureté d'un matériau survienne sur sa structure moléculaire, cela n'empêche pas que la dureté d'un objet puisse avoir un rôle pertinent dans l'explication d'un événement. Il y a toutefois une différence importante entre le mental et ces autres propriétés survenantes. En effet, alors que le mental nous apparaît résolument irréductible, ce n'est pas le cas pour les propriétés biologiques, géologiques et les autres propriétés survenantes de ce genre. Car c'est le postulat du réalisme du mental et l'existence du caractère subjectif et 'intrinsèque' des états mentaux qui s'opposent à la réduction, alors qu'il n'existe pas de postulat équivalent à celui du réalisme du mental ni d'aspect 'intrinsèque' aux propriétés biologiques ou géologiques. Au contraire, il y a une certaine plausibilité intuitive à supposer que des propriétés considérées comme physiques au

sens large du terme soient réductibles à des propriétés qui relèvent de la physique (en un sens plus restreint), comme par exemple certaines propriétés des molécules.

L'une des caractéristiques de l'argument de Kim qui le rend difficilement contestable est qu'il repose sur peu de présupposés. L'un de ces présupposés est le principe de l'exclusion causale, mais son rejet conduirait à trop de problèmes. Les deux autres présupposés sur lesquels est basé l'argument constituent l'essence même du physicalisme non-réductif, soit les thèses du physicalisme et de l'anti-réductionnisme. Le rejet du physicalisme conduirait au retour du dualisme des substances, ce qui ramènerait tous les problèmes liés à cette position, dont le problème de la causalité mentale (!). Reste donc le rejet de l'autre thèse fondamentale du physicalisme non-réductif: l'irréductibilité du mental au physique. En effet, c'est l'anti-réductionnisme qui nous oblige au moment critique à conclure (viii) (c'est-à-dire que P cause P^* et que M et M^* surviennent respectivement sur P et P^*), conduisant ainsi à l'épiphénoménisme. C'est pourquoi selon Kim il est possible échapper à l'argument de la survenance en acceptant la réduction de M à P et de M^* à P^*. Kim écrit:

> Car il est clair que le postulat implicite qui permet à l'argument de la survenance d'être concluant est une position anti-réductionniste en ce qui a trait au rapport corps-esprit; si les propriétés mentales sont conçues comme étant réductibles aux propriétés physiques d'une manière appropriée, on peut s'attendre à être à même de désamorcer l'argument (bien qu'évidemment les détails auraient besoin d'être travaillés).[29]

Mais que signifie l'adoption d'une approche réductionniste? Peut-on espérer rendre compte de la causalité mentale avec une telle approche? Pour que l'idée de causalité mentale ait un sens, il faut au minimum que les propriétés mentales existent et qu'elles possèdent des pouvoirs causaux; bref, que la thèse du réalisme du mental soit respectée. Mais il est impossible de réaliser cela avec le réductionnisme puisqu'une telle approche ne préserve pas l'existence des propriétés mentales; il est alors impossible de leur attribuer des pouvoirs causaux. Ainsi, si dans le cadre d'une réduction psychophysique un événement 'mental' m est

---

[29] *"For it is clear that the tacit assumption that gets the supervenience argument going is mind-body antireductionism; if the mental properties are viewed as reducible to physical properties in an appropriate way, we should expect to be able to disarm the argument (although of course the details will need to be worked out)."* Kim (1998), p. 46.

mentionné comme étant la cause d'un autre événement *e*, cet événement *m* n'est pas la cause de *e en vertu du fait* qu'il est un événement *mental* mais plutôt en vertu du fait qu'il est un événement *physique*. En effet, le réductionnisme implique l'abandon du *dualisme des propriétés*. En cela, le réductionnisme n'est guère différent de l'éliminativisme: dans les deux cas nous n'avons plus qu'un seul type de propriétés, des propriétés physiques. Lorsque l'on réduit des propriétés à d'autres propriétés, on effectue en fait une *identification* entre ces deux groupes de propriétés. En étant réduites aux propriétés physiques, les propriétés mentales sont donc identifiées aux propriétés physiques. Réduction et identification sont intimement liées, mais à proprement parler la différence entre la réduction et l'identification réside en ce que la première concerne les explications alors que la seconde s'applique aux objets. La réduction étant un processus épistémologique, la *théorie* des événements mentaux ('la psychologie populaire') qui est réduite se trouve à être *conservée* par la théorie de base, puisqu'elle en est *dérivable*. Nous pourrions alors toujours utiliser des expressions référant à nos propriétés mentales, mais la dénotation de ces expressions serait différente puisqu'elle référerait en fait à des propriétés physiques. Reste qu'ontologiquement parlant, le mental n'existe plus. Il ne faut donc pas s'étonner que le réalisme du mental s'oppose au réductionnisme. Bref, tant avec l'épiphénoménisme qu'avec le réductionnisme, il faut abandonner l'idée de la causalité mentale.

Il semble donc que peu importe l'option choisie, il soit impossible de rendre compte de la causalité mentale. Face à un tel constat, certains philosophes sont portés à mettre de côté les arguments métaphysiques pour retourner à leurs intuitions: n'apparaît-il pas évident que nos états mentaux existent et qu'ils sont souvent la cause de nos actions? Notre conviction en la capacité causale de nos états mentaux ne surpasse-elle pas notre attachement à n'importe quel principe métaphysique? Certes, mais comme je l'ai dit dès l'introduction, la question en litige n'est pas de savoir *si* nos états mentaux agissent causalement sur nos comportements, mais *comment* ils le font. Pour la question du 'si', nos convictions peuvent suffire, mais pour la question du 'comment', les arguments métaphysiques sont inévitables. Mais est-ce que l'argument de la survenance ne vient pas remettre en cause notre intuition quant à la possibilité

de la causalité mentale? Il semble que pour Kim ce ne sont pas tant nos intuitions concernant le 'si' que nos intuitions concernant le 'comment' qui soient remises en cause. Il écrit:

> Je crois que le problème plonge très profondément dans nos conceptions métaphysiques fondamentales quant à nous-mêmes et le monde dans lequel nous vivons et que nous devons y faire des ajustement assez drastiques si nous voulons sérieusement régler le problème. Lorsque ce problème sera réglé, notre métaphysique de l'esprit aura subi de sérieuses transformations. Il n'y a pas plus de dîners gratuits en philosophie de l'esprit que dans la vie, et je crois que les repas bon marché ne valent pas le coup. Nous serions aussi bien de prendre ce qu'il y a de mieux et d'en payer le prix. Lorsque le dualisme des substances a été confronté au problème de la causalité mentale, le dualisme fut le perdant: la substance mentale n'est plus avec nous. L'histoire pourrait très bien se répéter: dans la confrontation entre le dualisme des propriétés et la causalité mentale, le dualisme pourrait perdre encore une fois, laissant les propriétés mentales irréductibles dans la poussière.[30]

L'opinion de Kim est donc que notre conception métaphysique actuelle quant à la causalité mentale, le physicalisme non-réductif, n'est pas adéquate. C'est d'ailleurs là la conclusion de son argument de la survenance. Quant aux transformations que notre métaphysique doit alors subir, Kim nous indique la voie qu'il privilégie: l'abandon du dualisme des propriétés. Des deux options laissées ouvertes par l'argument de la survenance, l'épiphénoménisme et le réductionnisme, Kim choisit le réductionnisme. Il ne s'agit pourtant pas d'une approche ayant la faveur populaire. Malgré tout, Kim va aller jusqu'à proposer *deux* modèles réductifs qu'il est maintenant temps de présenter.

---

[30] "*I believe that the problem goes deep, deep into our fundamental metaphysical views about ourselves and the world we live in, and that we need to make fairly drastic adjustments if we are serious about coming to terms with the problem. When we are properly done with the problem, our metaphysics of the mind would have undergone some serious alterations. There are no free lunches in philosophy any more than in real life, and I believe the cheap ones aren't worth the money. We might as well go for the real stuff and pay the price. When substance dualism was confronted by the problem of mental causation, dualism was the loser: mental substance is no longer with us. History may very well repeat itself: in the confrontation between property dualism and mental causation, dualism may again lose out, leaving irreducible mental properties in the dust.*" Kim (1998), p. 60.

Chapitre 3: Les réductionnismes de Kim

Kim a présenté deux modèles réductifs destinés à servir d'alternatives aux approches du physicalisme non-réductif: il s'agit des modèles de *l'identification disjonctive* et de la *réduction fonctionnelle*. Est-ce à dire que Kim néglige les arguments ayant conduit à la chute du réductionnisme à la fin des années soixante? En effet, ce n'est pas parce que l'argument de la survenance épargne le réductionnisme que cette approche est nécessairement adéquate pour le problème corps-esprit. Lors de ma présentation du réductionnisme, j'ai mentionné les trois arguments principaux contre le réductionnisme classique, à savoir l'irréductibilité des qualia, l'anomalisme du mental et la réalisation multiple. De ces trois arguments, Kim ne s'attarde guère au premier lorsqu'il parle de réduction, mais il lui accorde plus d'attention lorsqu'il parle de définition fonctionnelle. J'y reviendrai dans la dernière section de ce chapitre et dans le chapitre 4. Quant au second argument, celui de Davidson, Kim l'écarte rapidement en déclarant que cet argument s'oppose au réalisme du mental puisqu'il va à l'encontre de l'existence de lois psychophysiques. L'idée est que s'il n'y a pas de lois psychophysiques, alors les états mentaux n'ont aucun rôle à jouer dans la détermination de nos comportements. Kim écrit:

> Car le monisme anomal implique ceci: *dans le monde selon Davidson, un réseau de relations causales absolument identique serait obtenu si on redistribuait les propriétés mentales des événements de n'importe quelle façon; on ne changerait pas la moindre relation causale si on assignait les propriétés mentales aux événements de façon arbitraire et au gré du hasard ou même si on enlevait complètement le mental du monde.* Le fait est que d'après le monisme anomal de Davidson, le mental ne joue aucun rôle causal.[31]

Le mental ne jouant aucun rôle causal, l'anomalisme du mental serait ainsi au mieux une forme d'épiphénoménisme. Kim considère ainsi qu'aucune théorie ne peut concilier l'anomalisme

---

[31] "*For anomalous monism entails this:* the very same network of causal relations would obtain in Davidson's world if you were to redistribute mental properties over its events any way you like; you would not disturb a single causal relation if you randomly and arbitrarily reassigned mental properties to events, or even removed mentality entirely from the world. *The fact is that under Davidson's anomalous monism, mentality does no causal work.*" Kim (1989b), aussi dans Kim (1993), essai 14, p. 269. Kim souligne.

du mental et la possibilité de la causalité mentale. Reste donc l'argument de la réalisation multiple, que Kim considére beaucoup plus sérieusement. En effet, en élaborant ses modèles réductifs, il semble s'être surtout efforcé de les rendre conformes à la réalisation multiple, cet objectif influençant directement les stratégies employées, comme nous allons le voir.

Je vais présenter les deux modèles réductifs de Kim et certaines critiques s'y rattachant. Nous verrons que le premier modèle, l'identification disjonctive, ne diffère du réductionnisme traditionnel que par sa portée plus limitée: Kim propose de faire des réductions locales, spécifiques à chaque espèce, puis d'en faire la disjonction afin d'obtenir un 'type' physique pouvant être identifié à l'état mental type à expliquer. Quant au second modèle, la réduction fonctionnelle, il a comme caractéristique d'être basé sur le fonctionnalisme: Kim fait valoir que les états fonctionnels tirent tous leurs pouvoirs causaux de leurs réalisateurs physiques.

Premier modèle réductif: L'identification disjonctive

Le modèle de l'identification disjonctive se propose naturellement en réponse à l'argument de la réalisation multiple. En supposant que l'identité des types n'a pas d'autre défaut que celui de ne pas s'appliquer à tous les organismes (autrement dit en l'absence d'une structure physique présente chez tous les individus de toutes les espèces présentant un état mental donné pouvant servir à l'identification avec cet état mental), il suffit alors de soutenir que le réductionnisme vaut *pour chaque espèce prise séparément*. En fait, il semble même que la thèse de la réalisation multiple *implique* l'existence de lois psychophysiques entre un état mental donné et son réalisateur physique particulier chez un être donné. En effet, s'il est vrai qu'en vertu de la réalisation multiple un état mental M peut être réalisé par plusieurs réalisateurs R_i différents (par hypothèse un R différent pour chaque espèce) et qu'il ne peut alors exister de relation *nécessaire* entre R_1 et M, ou R_2 et M, ou R_3 et M, etc., il peut néanmoins exister une relation *suffisante* entre un réalisateur donné R_i et M. Ainsi, pour tout état mental M, il existe une relation unidirectionnelle $R_i \rightarrow M$ garantie par la réalisation multiple de M. Bref, comme le dit Kim: "l'argument de la réalisation multiple montre peut-être que l'on ne

peut pas obtenir de fortes connections entre les propriétés mentales et les propriétés physiques; toutefois il *présuppose* que de *fortes connections spécifiques aux espèces* existent bien."[32]

D'après Kim, ce fait permet une résurrection partielle du réductionnisme. Évidemment, ces 'fortes connections' ou lois spécifiques à chaque espèce ne permettent pas d'effectuer une réduction de chaque état mental à une base physique et/ou biologique identique pour n'importe quel organisme, mais elles sont suffisantes pour l'obtention d'une série de réductions particulières pour chaque espèce, ou 'réductions locales'. Par exemple, le réalisateur de la douleur chez l'homme pourrait être *toujours* l'activation de fibres-C, chez le calmar cela pourrait être *toujours* l'activation de fibres-X, chez un extraterrestre *toujours* un mécanisme Y qui nous est inconnu et ainsi de suite pour tous les êtres doués de conscience. Il est alors impossible d'identifier la douleur avec l'activation de fibres-C puisque la douleur peut également être réalisée par l'activation de fibres-X, d'un mécanisme Y, etc. Par contre, on peut dire que *chez l'humain* la douleur est identique à l'activation des fibres-C, *chez le calmar* à l'activation des fibres-X, *chez l'extraterrestre* à l'activation d'un mécanisme Y, etc. Cette stratégie correspond très bien au réductionnisme, dont elle ne diffère que par la portée des réductions effectuées. Une théorie des états mentaux et une théorie des états physiques (une théorie pour chaque réalisateur), de même que des 'lois-ponts' sont toujours requises pour procéder à la réduction. On peut effectivement s'attendre à ce que les progrès de la science aillent dans le sens d'établir de plus en plus de corrélations entre des états mentaux et des états physiques particuliers pour chaque espèce. Ces corrélations pourraient servir de base à l'obtention de 'lois-ponts', conduisant à l'obtention de réductions locales. De telles réductions locales sont compatibles avec la réalisation multiple puisque pour chaque réalisateur différent qui pourrait se présenter, une réduction locale différente peut (en principe) s'appliquer.

---

[32] *"the multiple realization argument perhaps shows that the strong connectibility of mental properties vis-à-vis physical properties does not obtain; however, it* presupposes *that* species-specific strong connectibility *does hold."* Kim (1989b), aussi dans Kim (1993), essai 14, p. 274. Kim souligne.

Problèmes avec la réduction locale

Toutefois, le recours aux réductions locales ne permet pas d'échapper à tous les problèmes. (Tout comme dans la présentation du réductionnisme au chapitre 1, je vais m'en tenir à la version nagelien de la réduction qui est la plus largement acceptée et que Kim utilise pour cette raison.) Premièrement, plusieurs philosophes ont objecté que la réalisation multiple pourrait bien aller plus loin que les espèces biologiques, c'est-à-dire qu'à l'intérieur d'une même espèce différents réalisateurs pourraient être responsables de l'apparition d'un même état mental, et que même chez un individu particulier, le réalisateur d'un état mental pourrait changer avec le temps étant donné la croissance puis le vieillissement ou des blessures au cerveau. En réponse à cette objection, on peut remarquer que la psychologie humaine présente une certaine uniformité (à preuve l'existence de la psychologie en tant que discipline). Or, si la psychologie est effectivement réalisée physiquement, alors à cette uniformité de la psychologie doit correspondre une uniformité physiologique chez les humains. Mais cette réponse n'est absolument pas satisfaisante: l'argument de la réalisation multiple est justement là pour dire qu'une même psychologie peut être réalisée par des systèmes hautement hétéroclites. On peut toutefois raisonnablement supposer qu'il existe une certaine uniformité dans la neurologie humaine, de même que dans celle de chacune des autres espèces. Reste que, dans le pire des cas, on peut se retrouver avec des 'lois-ponts' différentes pour chaque individu et pour chaque moment particulier de la vie de cet individu, rendant alors impossible d'effectuer des réductions locales utiles (rendu à ce point, cependant, on peut difficilement parler de 'lois', puisque ces supposées lois ne couvrent plus qu'un seul événement). En fait, poussée à l'extrême cette situation me semble revenir à l'anomalisme du mental: seules les occurrences d'événements mentaux et d'événements physiques peuvent être identifiées.

Le deuxième problème est plus grave que le premier. Car si on peut encore raisonnablement supposer qu'il existe des 'lois-ponts' suffisamment stables à travers le temps et applicables à suffisamment d'individus pour produire des réductions locales, on peut se demander ce qui *explique l'existence de ces 'lois-ponts'*. En effet, d'après le modèle réductif de Nagel, les relations entre la théorie réduite et la théorie de base sont découvertes empiriquement et

doivent être acceptées comme un fait brut. La réduction ne peut alors répondre à des questions comme "pourquoi la sensation de picotement est-elle réalisée par l'activation de tel groupe de neurones?" ou "pourquoi n'est-ce pas la sensation de douleur ou même la sensation de rouge qui est réalisée par ce groupe de neurones?". Bref, tout comme la survenance ne s'explique pas par elle-même, la réduction nagelienne ne propose aucune justification de son application, ce qui fait que la réduction ne constitue pas une véritable explication des états mentaux: "Car c'est l'explication de ces 'lois-ponts', l'explication du pourquoi de l'existence de ces corrélations psychophysiques particulières qui est au coeur de la demande d'une explication du mental."[33] La réduction du mental au physique créerait donc toute une classe de phénomènes inexpliqués.

Mais la réduction ne conduit-elle pas à l'identification, et par le fait même ne donne-t-elle pas une explication de la relation entre le mental et le physique? C'est effectivement là l'opinion commune, d'où le troisième problème: la réduction nagelienne n'implique pas l'identification. En effet, passer de la réduction nagelienne à l'identification, c'est faire un saut non-justifié. Car tout ce que Nagel demande pour effectuer la réduction d'une théorie à une autre, c'est que la première soit dérivable de l'autre. Or, aucune identification n'est nécessaire pour cela: une forme de dépendance entre les états mentaux et les états physiques suffit. Pourtant, nous sommes en droit d'attendre d'une réduction qu'elle conduise à une simplification ontologique. Grâce à l'identification, il est possible d'obtenir une telle simplification fournissant du même coup l'explication recherchée pour les 'lois-ponts': la propriété mentale M et la propriété physique P sont en corrélation parce que $M = P$ et qu'il n'y a donc qu'une seule propriété, plutôt que deux qui corréleraient de façon inexpliquée. Jean Chrétien est présent chaque fois que le premier ministre du Canada est présent parce que Jean Chrétien *est* le premier ministre du Canada. Pour obtenir une simplification ontologique à partir de la réduction nagelienne, il nous faudrait donc trouver un moyen de transformer les corrélations en identités.

---

[33] "*For it is the explanation of these bridge laws, an explanation of why there are just these mind-body correlations, that is at the heart of the demand for an explanation of mentality.*" Kim (1998), p. 96.

Enfin, même si nous obtenions une identité, un autre problème se présenterait. Car la propriété mentale M est supposée être la même pour plusieurs espèces, alors que les réaliseurs de M diffèrent selon les espèces. Mais dire qu'une propriété mentale M est identique à un réalisateur R_1 pour les humains et à un réalisateur R_2 pour les calmars, etc., contrevient à la loi de l'identité des indiscernables selon laquelle, pour être identiques, deux choses doivent être indiscernables. En effet, si R_1 et R_2 ne sont pas identiques, comment M pourrait-il leur être identique à tous les deux?

La stratégie disjonctive appliquée aux réductions locales

Une solution est de dire que M est identique avec la *disjonction* de tous ses réalisateurs possibles. À première vue, cette approche semble intéressante: puisque la propriété mentale M est nécessairement réalisée par l'un ou l'autre de ses réalisateurs R_i, une identité nécessaire pourrait ainsi être affirmée comme suit: $M = R_1 \vee R_2 \vee ... \vee R_n$. Avec une telle identité, nous obtenons une simplicité ontologique: M n'est rien de plus que l'un ou l'autre de ses réalisateurs. Nous obtenons également une explication de la relation entre M et les R_i: ils corrèlent car ils sont identiques. Quant au premier problème, il est moins important car bien qu'on ne puisse montrer que c'est bien le cas, on peut néanmoins soutenir que les individus d'une même espèce partagent *grosso modo* la même neurologie. Tel est donc le modèle de l'identification disjonctive de Kim.

Toutefois, quel est le statut d'une telle propriété disjonctive? Il n'existe certainement rien d'ontologiquement disjonctif, c'est-à-dire quelque chose qui posséderait non pas une *propriété-X* ou une *propriété-Y*, mais qui posséderait une *propriété-X-ou-Y*, comme par exemple un objet qui soit rouge-ou-bleu. De plus, il est difficile de voir en quoi une propriété disjonctive serait utile pour formuler des explications. Par exemple, supposons qu'un meurtre a été commis par le colonel Mustard ou le professeur Plum. Peut-on dire alors que nous connaissons le coupable? En disant que le meurtrier est le colonel Mustard ou le professeur Plum, nous ne mentionnons pas un *unique suspect disjonctif*, mais exprimons une *disjonction de deux suspects différents*. Nous savons que l'un ou l'autre de ces deux suspects est le coupable, mais nous ne savons pas lequel; le crime n'est pas résolu tant que nous n'avons pas trouvé lequel

des deux est le coupable. Autre exemple: un incendie est dû à une défaillance électrique ou à un pyromane. Là encore, nous avons le choix entre deux causes plutôt que d'avoir une seule cause disjonctive. La disjonction de propriétés hétérogènes ne crée donc pas de nouvelles propriétés pouvant être employées comme cause ou explication.

Kim soutient également que de telles propriétés disjonctives ne peuvent être utilement employées dans des lois. On peut tenter de confirmer cette affirmation en tentant de formuler des 'lois disjonctives', c'est-à-dire des lois comprenant une cause ou une explication disjonctive. Par exemple, prenons la loi disjonctive suivante: les défaillances électriques, *D*, ou les pyromanes, *P*, causent des incendies, *I*. Kim formaliserait cet exemple comme suit:

$$(D \vee P) \rightarrow I$$

L'argument de Kim consiste à dire qu'il est possible de vérifier cette loi empiriquement, c'est-à-dire d'accumuler des cas où des incendies ont été causés par des défaillances électriques ou des pyromanes, sans que cette vérification vaille nécessairement pour des lois logiquement équivalentes. En effet, notre loi disjonctive est logiquement équivalente à la *conjonction* des lois "les défaillances électriques causent des incendies" et "les pyromanes causent des incendies". Bref:

$$((D \vee P) \rightarrow I) \equiv ((D \rightarrow I) \wedge (P \rightarrow I))$$

Or, si tous les cas observés sont des cas de défaillances électriques, seule la partie "les défaillances électriques causent des incendies"est confirmée, non l'autre. Il apparaît donc possible de confirmer une loi disjonctive sans que cette confirmation vaille pour des lois logiquement équivalentes.

Il y a toutefois un problème de taille avec cette démonstration. En effet, une formalisation adéquate d'une loi causale ne peut pas utiliser la relation d'implication logique (→); il faut obligatoirement utiliser l'implication *causale* ($\rightarrow_c$). Une formalisation adéquate de la loi disjonctive "les défaillances électriques ou les pyromanes causent des incendies" est donc:

$$((D \vee P) \rightarrow_c I)$$

et une formulation adéquate de la conjonction des lois "les défaillances électriques causent des incendies" et "les pyromanes causent des incendies" s'écrit:

$((D \rightarrow_c I) \wedge (P \rightarrow_c I))$

Le véritable problème pour Kim est qu'il est alors faux de dire que $((D \vee P) \rightarrow_c I)$ est logiquement équivalent à $((D \rightarrow_c I) \wedge (P \rightarrow_c I))$[34]. En effet, l'implication causale n'a pas les mêmes propriétés que l'implication logique. Kim n'est donc pas justifié de dire que les propriétés disjonctives ne peuvent être utilement employées dans des lois, du moins pas sur la base de cet argument.

Malgré que ce dernier argument ne soit pas valide, la stratégie de l'identification disjonctive se révèle tout de même moins intéressante que ce qu'elle pouvait apparaître à première vue, puisqu'elle aboutit à l'emploi de propriétés qui ne peuvent être utilement employées comme cause ou comme explication et dont le statut ontologique est pour le moins discutable. Or, sans l'identification disjonctive, la réduction locale n'est pas capable de fournir un modèle réductif viable. C'est pour ces raisons que Kim a fini par rejeter cette approche et qu'il a alors proposé un second modèle réductif: la réduction fonctionnelle.

Second modèle réductif: La réduction fonctionnelle

Avec sa réduction fonctionnelle, Kim cherche encore à proposer un modèle qui soit en accord avec la thèse de la réalisation multiple tout en offrant une explication de la survenance du mental sur le physique, cette explication devant (nécessairement, semble-t-il) passer par l'identification. Cette réduction fonctionnelle est basée sur la thèse du *fonctionnalisme physicaliste* ('*physicalist fonctionalism*' ou '*physical realisationism*'), qui consiste à dire que les propriétés mentales, si elles sont réalisées, doivent nécessairement être réalisées physiquement, c'est-à-dire qu'aucune propriété mentale ne peut avoir de réalisateur non-physique. À cet égard, cette approche de Kim ne diffère pas du courant principal du

[34] Je dois cette observation à J. Nicolas Kaufmann.

fonctionnalisme dont j'ai parlé dans la section sur le physicalisme non-réductif. Kim décrit simplement ici une version résolument physicaliste du fonctionnalisme:

> Cette thèse est ainsi l'équivalent de la conjonction du physicalisme et de la conception fonctionnaliste des états mentaux, aussi "fonctionnalisme physicaliste" serait un autre nom approprié à cette position . Le fonctionnalisme considère les propriétés mentales comme des propriétés fonctionnelles, des propriétés définies par leurs rôles en tant qu'intermédiaires causals entre les sensations et le comportement, et la version physicaliste du fonctionnalisme considère les propriétés physiques comme les seuls occupants ou réalisateurs potentiels de ces rôles causaux.[35]

La réduction fonctionnelle se veut très près du fonctionnalisme, tout comme la réduction locale était très semblable au modèle du réductionnisme classique. Les propriétés mentales sont ainsi définies par leur rôle causal, ce rôle étant rempli par leur réalisateur physique. Être dans un état mental, c'est donc être dans un état ayant typiquement telles et telles causes et tels et tels effets. Le fait pour une propriété physique de remplir un rôle spécifique, c'est-à-dire d'être ou non le réalisateur d'une propriété fonctionnelle, dépend alors des relations causales entretenues avec les autres propriétés. Puisque seules importent les relations causales des propriétés, les caractéristiques structurelles ou compositionnelles des réalisateurs n'ont pas d'importance. C'est donc dire que ce modèle de Kim, à l'image du fonctionnalisme, respecte la thèse de la réalisation multiple.

Selon Kim, le fonctionnalisme implique que les propriétés fonctionnelles sont des propriétés de *second ordre*: pour un individu x, avoir une propriété fonctionnelle F consiste en la possession par x d'une propriété de premier ordre P satisfaisant une condition C, cette condition spécifiant les causes et effets typiques de F. De façon générale, voici comment Kim définit une propriété de second ordre:

---

[35] *"The thesis therefore is equivalent to the conjunction of physicalism with the functionalist conception of mental properties, and "physicalist functionalism" would be an equally good name for this position. Functionalism takes mental properties and kinds as functional properties, properties specified in terms of their roles as causal intermediaries between sensory inputs and behavioral outputs, and the physicalist form of functionalism takes physical properties as the only potential occupants, or "realizers", of these causal roles."* Kim (1998), p. 19.

Propriété de second ordre: *F* est une *propriété de second ordre* par rapport à un ensemble de propriétés de premier ordre **B** si et seulement si *F* est la propriété d'avoir une propriété *P* de l'ensemble **B** telle que *C*(*P*), où *C* spécifie une condition s'appliquant aux membres de **B**.[36]

Les propriétés de second ordre seraient donc des propriétés générées par quantification existentielle sur des propriétés de premier ordre, les propriétés de premier ordre qui satisfont la condition *C* étant les réalisateurs des propriétés de second ordre. Les propriétés fonctionnelles seraient ainsi un cas spécial de propriétés de second ordre, c'est-à-dire les propriétés de second ordre dont la condition *C* concerne les relations causales entre les propriétés de premier ordre. Enfin, il faut préciser qu'une propriété donnée n'est pas une propriété de premier ordre de façon absolue: elle peut être une propriété de second ordre par rapport à d'autres propriétés. Dans le cas des propriétés mentales, les propriétés de base, ou de premier ordre, consistent bien sûr en des propriétés non-mentales, c'est-à-dire des propriétés physico-chimiques, biologiques ou béhaviorales.

Compte tenu de cette caractérisation des propriétés fonctionnelles, on peut dire que pour un individu donné la détermination de ses caractéristiques physiques détermine s'il possède une propriété mentale donnée. En effet, si la présence d'une propriété mentale *M* dépend de la présence de son réalisateur physique *P*, alors la présence de *P* (définie exclusivement en termes de relations causales) détermine la présence de *M*. En fait, si *P* réalise *M* dans un système *s*, alors *P* réalise *M* dans tous les systèmes causalement identiques à *s* et soumis aux mêmes lois de la nature, c'est-à-dire dans des mondes nomologiquement identiques. *P* est donc nomologiquement suffisant pour *M*, ce qui signifie que la thèse du fonctionnalisme physicaliste implique la survenance forte (nomologique) de *M* sur *P*. La force de la relation de survenance est limitée à celle de la nécessité nomologique car *P* réalise *M* en vertu de ses relations causales avec les autres occurrences de propriétés; si les lois de la nature sont changées, *P* pourrait ne plus avoir les mêmes relations causales avec elles.

---

[36] "F *is a second-order property over set* ***B*** *of base (or first-order) properties iff* F *is the property of having some property* P *in* ***B*** *such that* D(P), *where* D *specifies a condition on members of* ***B***." Kim (1998), p. 20.

Le principe de l'héritage des pouvoirs causaux

Cependant, la réalisation serait une relation plus forte que la survenance, menant à *l'identification des pouvoirs causaux* des états mentaux avec ceux de leurs réalisateurs. En effet, puisque pour un individu *x*, *avoir une propriété fonctionnelle F, c'est avoir un des réalisateurs physiques de F*, il semble que le fonctionnalisme physicaliste implique une forme de *réduction fonctionnelle*. Ainsi, une propriété fonctionnelle ne posséderait aucun pouvoir causal 'en elle-même', si on peut s'exprimer ainsi; tous les pouvoirs causaux qu'impliquerait une propriété fonctionnelle seraient en fait possédés par son réalisateur. Kim parle alors du "principe de l'héritage des pouvoirs causaux":

> **Principe de l'héritage des pouvoirs causaux:** Si une propriété de second ordre *F* est réalisée à un moment donné par une propriété de premier ordre *H* (c'est-à-dire que si *F* est présent à un moment donné en vertu du fait que l'un de ses réalisateurs, *H*, est présent à ce moment), alors les pouvoirs causaux de cette instance de *F* sont identiques aux pouvoirs causaux (ou sont un sous-ensemble des pouvoirs causaux) de *H* (ou de cette instance de *H*).[37]

Le principe de l'héritage des pouvoirs causaux affirme que les pouvoirs causaux des états mentaux sont identiques à ceux de leurs réalisateurs. D'après Kim, cette identification des pouvoirs causaux permet à son approche de la réduction fonctionnelle d'expliquer la survenance psychophysique: le mental survient sur le physique parce que les propriétés mentales sont des propriétés fonctionnelles de second ordre ayant des réalisateurs physiques (et jamais de réalisateurs non-physiques). De plus, *M* est présent chaque fois que *P* est réalisé par un système *s* parce qu'avoir *M* c'est par définition avoir une propriété ayant une rôle causal *C* et que, dans les systèmes comme *s*, *P* est la propriété (ou l'une des propriétés) satisfaisant la condition *C*. Ainsi, pour les systèmes comme *s*, avoir *M* c'est avoir *P*. Kim est très clair sur la connotation identificatrice et réductive de cette affirmation:

---

[37] *"If a second-order property* F *is realized on a given occasion by a first-order property* H *(that is, if* F *is instantiated on a given occasion in virtue of the fact that one of its realizers,* H*, is instantiated on that occasion), then the causal powers of this particular instance of* F *are identical with (or are a subset of) the causal powers of* H *(or of this instance of* H*)."* Kim (1998), p. 54.

> Il ne faut pas comprendre que la propriété *M* apparaît ou émerge magiquement lorsque des systèmes ont *P*. C'est plutôt qu'avoir *M* pour ces systèmes *est* simplement avoir *P*. Nous pourrions même dire, employant une expression familière mais défraîchie du réductionnisme, qu'avoir *M*, pour ces systèmes, n'est "rien de plus" qu'avoir *P*. [...] D'un point de vue métaphysique, par conséquent, l'idée que les propriétés mentales sont réalisées par des propriétés physiques va beaucoup plus loin que des idées du genre que les propriétés mentales ont des "corrélats physiques", des "substrats neuraux", ou des "bases physiques subvenantes", etc.; contrairement à la "réalisation", ces idées ne sont pas capables d'expliquer pourquoi une propriété mentale donnée corrèle ou apparaît avec certaines propriétés physiques, et elles ne permettent pas l'emploi d'expressions réductives comme: "Avoir *M*, pour les systèmes appropriés, consiste en, ou est simplement, avoir *P*."[38]

Il s'agit donc bien d'une forme de réductionnisme que nous présente Kim à partir de l'affirmation apparemment innocente d'un fonctionnalisme physicaliste. Kim soutient même que ce réductionnisme fonctionnel correspond beaucoup mieux à la manière dont la science procède que le modèle réductif de Nagel. Ainsi, en science, pour réduire une propriété ou un phénomène, il faut d'abord le caractériser relationnellement, c'est-à-dire en termes de ses relations causales avec les autres propriétés et phénomènes. Pour prendre un des exemples favoris de Kim, le gène est ce qui permet de transmettre certaines caractéristiques biologiques des parents aux enfants. Ou alors: un corps est transparent si la lumière peut passer au travers. Il suffit alors de trouver les mécanismes ou les propriétés physiques (souvent microphysiques) qui remplissent ces conditions. Comme il y a souvent plusieurs possibilités, la réalisation multiple et la dépendance aux lois physiques en vigueur dans le monde considéré s'obtiennent dans ces cas comme dans le cas des états mentaux. Le gène est ainsi une notion fonctionnelle

---

[38] *"It isn't that when certain systems instantiate* P, *mental property* M *magically emerges or supervenes (in the dictionary sense of "supervene"). It is rather that having* M *for these systems, simply is having* P. *We might even say, using a familiar if shopworn reductive idiom, that having* M, *for these systems, is "nothing over and above" having* P. *[...] From a metaphysical point of view, therefore, the idea that mental properties are realized by physical properties goes considerably beyond such ideas as that mental properties have "physical correlates" or "neural substrates", that they have "physical supervenience bases", and the like; unlike "realization", these ideas are not capable of generating an explanation of why a given mental property arises out of, or correlates with, certain physical properties, and do not warrant reductive talk like "Having* M, *for appropriate systems,* consists in, *or* just is, *having* P.*"* Kim (1998), p. 24.

réalisée par la molécule d'A.D.N.[39], mais pourrait possiblement être réalisé par un autre mécanisme chez un être radicalement différent ou dans un monde dont les lois de la nature seraient différentes. De même pour la transparence: différentes structures moléculaires expriment la propriété d'être perméable à la lumière (le verre, plusieurs sortes de plastiques, certains cristaux, certaines membranes biologiques, etc.), mais dans un monde soumis à des lois physiques différentes, d'autres structures que celles-là pourraient être transparentes.

Bref, le modèle de la réduction fonctionnelle consiste, en premier lieu, à fonctionnaliser la propriété *M* à réduire, c'est-à-dire à la définir relationnellement ou *extrinsèquement* (par opposition à *intrinsèquement*). En second lieu, une fois que la propriété *M* a été ramenée à la propriété d'avoir une propriété ayant tels et tels pouvoirs causaux, il faut trouver la propriété de premier ordre *P* qui corresponde exactement à cette caractérisation pour un système *s* donné; enfin, puisqu'en général la propriété d'avoir une propriété donnée est identique à cette même propriété, il s'ensuit que la propriété d'avoir la propriété *M* est identique à la propriété *M* elle-même et on peut conclure que, pour le système *s*, $M = P$[40]. Ainsi, selon ce modèle de la réduction fonctionnelle, le mental n'existe plus ontologiquement, mais seulement 'catégoriellement': les propriétés mentales n'existent plus puisqu'elles sont identifiées avec leurs réalisateurs. Ce modèle réductif s'oppose donc au dualisme des propriétés malgré qu'il s'inspire du fonctionnalisme qui, lui, était compatible avec le dualisme des propriétés.

La réduction fonctionnelle et la réalisation multiple

Bien que, d'une part, l'identification de *M* et *P* soit compréhensible en l'absence de réalisation multiple et qu'il soit clair, d'autre part, que le fonctionnalisme permet la réalisation multiple, je crois qu'il n'est pas inutile de revenir plus en détails sur la manière dont Kim effectue une identification entre *M* et *P* lorsque la propriété fonctionnelle *M* possède plus d'un réalisateur. La stratégie de Kim repose de façon importante sur l'idée que les propriétés fonctionnelles sont

---

[39] Plus précisément, les diverses fonctions du gène sont remplies par diverses parties de la molécule d'A.D.N., comme me l'a souligné J. Nicolas Kaufmann. Reste que l'ensemble des fonctions attribuées au gène sont réalisées par l'ensemble de la molécule d'A.D.N.

[40] Voir Kim (1998), p. 98.

des propriétés de second ordre. En effet, Kim fait valoir que les propriétés mentales n'existent pas *en soi*, c'est-à-dire *en plus* des propriétés physiques qui les réalisent.

Pour étayer son point, Kim commence par soutenir qu'effectuer une quantification existentielle sur un domaine de propriétés ne fait pas *vraiment* apparaître un nouvel ensemble de propriétés, que si par une simple opération logique nous pouvions changer notre ontologie, cela serait "de la pure magie". Dans une perspective où des propriétés distinctes doivent représenter des pouvoirs causaux distincts, comme Kim semble le proposer, cela serait effectivement le cas. Il a été dit qu'avoir une propriété de second ordre M revenait à avoir une propriété de premier ordre satisfaisant une caractérisation C. S'il existe deux propriétés de premier ordre satisfaisant C, disons P_1 et P_2, alors avoir M c'est avoir P_1 ou avoir P_2. Le fait d'avoir M revient au *fait disjonctif* (ou *proposition disjonctive*) d'avoir l'un ou l'autre des réalisateurs. Toutefois, dire qu'avoir M = avoir P_1 ou avoir P_2 n'implique pas que la propriété M = la propriété disjonctive $P_1 \vee P_2$, comme nous l'avons vu dans la section sur l'identification disjonctive. En disant que le meurtrier est le colonel Mustard ou le professeur Plum, nous affirmons la disjonction de deux suspects différents, non la présence d'un unique suspect disjonctif. Dans la plupart des cas, le "ou" qui semble créer un prédicat disjonctif dans ce genre de phrase est en fait une abréviation pour la disjonction de deux affirmations. Bref, "il n'y a aucun besoin de considérer M comme une propriété de plein droit - pas même une propriété disjonctive ayant les Ps comme constituants. En quantifiant sur des propriétés, nous ne pouvons pas plus créer de nouvelles propriétés qu'en quantifiant sur des individus nous ne pouvons créer de nouveaux individus."[41]

C'est pourquoi Kim propose que plutôt que de parler de 'propriétés de second ordre', nous devrions parler de 'descriptions', 'd'appellations' ou de 'concepts' de second ordre. De tels concepts de second ordre sont utiles lorsque nous sommes incapables ou ne voulons pas utiliser des concepts de premier ordre pour ce dont nous parlons. Par exemple, il peut être utile

---

[41] *"There is no need here to think of M itself as a property in its own right - not even a disjunctive property with the Ps as disjuncts. By quantifying over properties, we cannot create new properties any more than by quantifying over individuals we can create new individuals."* Kim (1998), p. 103-104.

d'appeler 'somnifère' deux médicaments différents obtenant leur effet de manières appréciablement différentes. Le concept de second ordre 'somnifère' peut ainsi être réalisé par deux substances chimiques différentes, S_1 et S_2, tenant lieu de propriétés de premier ordre. Kim propose ainsi une conception 'restrictive' des propriétés par opposition à une conception 'large' ou 'inclusive'. Selon lui, voir *M* comme une véritable propriété serait accepter l'idée d'une propriété disjonctive (par exemple $P_1 \vee P_2$), mais nous avons vu que cette stratégie comporte des inconvénients, comme le fait que les propriétés disjonctives ne peuvent être employées dans des lois de la même manière que les autres propriétés. Enfin, le fait de ne pas considérer *M* comme une véritable propriété permet déjà de répondre à un problème potentiel, soit le fait qu'identifier *M* et *P* revient à identifier une propriété de second ordre avec une propriété de premier ordre, une propriété relationnelle avec une propriété intrinsèque, un rôle causal avec son occupant. Ce genre d'identification n'est pas nécessairement évident à première vue. Mais si on dit que seuls les *P*s existent, cette difficulté disparaît.

Revenons à la façon dont Kim explique l'identification de *M* avec un *P* donné dans un cas de réalisation multiple. La réduction fonctionnelle de *M* consiste à identifier *M* avec son réalisateur P_i approprié par rapport à une espèce ou à une structure donnée. Ainsi, *M est* P_1 pour l'espèce 1, P_2 pour l'espèce 2, P_3 pour l'espèce 3, etc. Étant donné que chaque instance de *M* possède exactement les mêmes pouvoirs causaux que son réalisateur en vertu du principe de l'héritage des pouvoirs causaux, tout le travail causal ou explicatif produit par une instance de *M* en vertu du fait qu'elle est réalisé par P_1 est effectué par P_1, et de même pour les autres instances de *M* et leurs réalisateurs. Chaque instance de *M* est une instance de P_1, ou de P_2, ou de P_3, etc. Puisque *M* n'est pas une propriété à proprement parler, il ne s'agit pas d'une identification disjonctive. C'est plutôt que *M* désigne la classe des objets ou propriétés ayant une certaine caractérisation fonctionnelle en commun, et que les *P*s sont justement ces objets ou propriétés. De plus, on voit facilement comment Kim peut éviter le problème de l'exclusion avec une telle approche: puisque chaque occurrence de *M* est en fait une occurrence de l'un ou l'autre des *P*s, il n'y a qu'un seul événement invoqué comme cause, pas deux. La compétition causale est ainsi impossible.

Dernières remarques sur la réduction fonctionnelle

La réduction fonctionnelle permet de réduire le mental au physique, ce qui pour Kim signifie la fin de nos ennuis avec la causalité mentale: le problème de l'exclusion est évité et les 'propriétés mentales' sont causalement efficaces dans la mesure où leurs réaliseurs le sont. Kim ajoute toutefois un bémol:

> Ceci résout le problème de l'efficacité causale des propriétés mentales fonctionnalisables. Ce sont les propriétés mentales qui résistent à la fonctionnalisation qui présentent des difficultés lorsque nous essayons de rendre compte de leurs pouvoirs causaux. Tant que nous croyons à l'existence possible de propriétés mentales non-fonctionnalisables, par exemple les qualia, qui néanmoins surviennent sur des propriétés physiques, nous sommes confrontés au problème de la causalité mentale.[42]

Que contrairement aux qualia les états mentaux intentionnels soient perçus comme aisément fonctionnalisables est une conviction largement partagée parmi les philosophes aujourd'hui. La raison en est que les états intentionnels consistent en des attitudes propositionnelles dont le contenu est déterminé par la signification des termes employés pour le représenter. Or, ce contenu est individué en fonction du monde que l'état intentionel représente. C'est donc dire que la signification d'un état intentionnel est extrinsèque. Et puisque fonctionnaliser c'est rendre extrinsèque un concept ou une propriété, on voit que les états mentaux intentionnels se prêtent bien à la fonctionnalisation. Par contre, il en est autrement des qualia qui sont réputés constitués les propriétés intrinsèques par excellence, puisque toute tentative visant à rendre les qualia extrinsèques les feraient disparaître. Il semble donc que le modèle de la réduction fonctionnelle de Kim achoppe sur le cas de la conscience phénoménale.

Quoi qu'il en soit, ce dernier défaut n'enlève pas nécessairement tout intérêt à l'approche de Kim. L'affirmation selon laquelle la réduction est essentiellement une fonctionnalisation pourrait bien être vraie. Il y a en effet une certaine plausibilité à ce que la science procède

---

[42] "*This solves the problem of causal efficacy for functionalizable mental properties. It is those mental properties that resist functionalization that present difficulties when we try to give an account of their causal powers. So long as we think there possibly are nonfunctionalizable mental properties, for example, qualia, which nonetheless supervene on the physical properties, we are faced with the problem of mental causation.*" Kim (1998), p. 116.

d'une manière similaire pour effectuer une simplification des phénomènes à étudier. De plus, ce modèle fonctionnel de la réduction apparaît plus crédible que celui de Nagel dont l'attention est portée sur la dérivation de lois. Premièrement, il y a de bonnes raisons de croire que la réduction fonctionnelle est particulièrement bien adaptée pour traiter des états mentaux intentionnels, comme on vient de le voir un peu plus haut. Deuxièmement, il s'agit d'un modèle qui permet la réalisation multiple, contrairement à celui de Nagel qui exige que chaque propriété mentale reçoive une propriété physique nomologiquement coextensive qui soit valide pour toutes les espèces et types de structure, ce qui fait du réductionnisme une cible facile. Enfin, alors que le modèle de Nagel n'implique pas nécessairement l'identité, la réduction fonctionnelle, elle, grâce au principe de l'héritage des pouvoirs causaux, affirme l'identité des 'propriétés' de second ordre et des propriétés de premier ordre. Cette identité permet d'expliquer pourquoi il y a une corrélation entre M et P. Kim affirme que l'identité entre M et P_1 est nomologiquement nécessaire, car si l'identité $M = P_1$ vaut pour un monde W_1, elle vaudra nécessairement pour tous les mondes ayant les mêmes lois physiques que W_1.

En terminant, on peut remarquer que si le modèle de la réduction fonctionnelle de Kim a quelque valeur, il conduit à ce que la réductibilité d'une propriété dépend de la possibilité de sa fonctionnalisation; autrement dit, si oui ou non cette propriété peut être conçue comme une propriété fonctionnelle de second ordre par rapport à une propriété de base. Les fameuses 'lois-ponts' de Nagel ne seraient alors plus nécessaires pour obtenir une réduction. Dans le cadre du problème corps-esprit, la question est ainsi de savoir si le mental peut être fonctionnalisé ou si, au contraire, il ne peut l'être par principe. Cela conduit Kim à une conclusion quelque peu étonnante:

> Si la conception fonctionnaliste du mental est correcte - correcte pour toutes les propriétés mentales - alors la réduction de l'esprit au corps est, sinon pratiquement réalisable, du moins possible en principe. Ceci est contraire à une affirmation de la sagesse philosophique contemporaine selon laquelle le fonctionnalisme, contrairement au physicalisme de l'identité des types, est une forme - en fait la principale forme contemporaine - d'anti-réductionnisme psychophysique. Ce que je propose ici est exactement le contraire - que la conception fonctionnaliste des propriétés mentales est

requise pour la réduction psychophysique. En fait elle est nécessaire et suffisante pour la réduction.[43]

Par conséquent, si le modèle de Kim est valide, autrement dit si la thèse du fonctionnalisme physicaliste et le principe de l'héritage des pouvoirs causaux sont vrais et si on peut conclure à la réduction des propriétés mentales aux propriétés physiques d'une manière satisfaisante, alors le fonctionnalisme et la réduction psychophysique vont rester ou tomber de concert; la survie de l'un ne se fera pas au détriment de l'autre. Et cela est vrai même si nous en arrivons à la conclusion que les qualia ne sont pas fonctionnalisables, comme la tendance actuelle semble l'indiquer et comme le croit Kim lui-même.

[43] "*If the functionalist conception of the mental is correct - correct for all mental properties - then mind-body reduction is in principle possible, if not practically feasible. This is contrary to one piece of current philosophical wisdom, the claim that functionalism, as distinguished from classic type physicalism, is a form - in fact the principal contemporary form - of mind-body antireductionism. What I am urging here is the exact opposite - that the functionalist conception of mental properties is* required *for mind-body reduction. In fact it is necessary and sufficient for reducibility.*" Kim (1998), p. 101. Kim souligne.

Chapitre 4: Les difficultés

Ce dernier chapitre est consacré aux difficultés qui persistent dans les propositions de Kim. Je vais commencer par remettre en question le présupposé de Kim selon lequel l'identité possède une vertu explicative de loin supérieure à celle de la corrélation, de la dépendance ou de la survenance. Puis je reviendrai sur le modèle de la réduction fonctionnelle de Kim pour discuter de l'efficacité causale du mental et de la place des qualia. Enfin je présenterai et discuterai une nouvelle voie que semble disposé à prendre Kim quant aux qualia, c'est-à-dire l'émergentisme.

De la portée explicative de l'identité

L'un des arguments les plus fréquents que Kim utilise pour montrer la supériorité de l'identification sur les autres types de relation, c'est la portée explicative. Ainsi, Kim soutient de façon convaincante qu'avec les relations de survenance et de dépendance, nous n'avons pas d'explication de cette relation. Par exemple, en affirmant qu'un état mental *M* survient sur un état physique *P*, nous ne disons rien sur la raison de cette survenance psychophysique: pourquoi *M* survient-il sur *P* plutôt que sur *P**? Ou pourquoi *M** ne survient-il pas sur *P*? C'est pourquoi Kim considère que l'affirmation de la dépendance ou de la survenance du mental sur le physique ne constitue que la formulation du problème corps-esprit, pas une solution à celui-ci. Par contre, d'après Kim, une fois qu'un événement mental et un événement physique sont identifiés, on ne se demande pas pourquoi tel événement mental et tel événement physique présentent une relation de survenance: *M* semble survenir sur *P* parce qu'en fait *M* et *P* sont un seul et même événement. La réduction de *M* à *P* impliquant une identification, la réduction présente ainsi un avantage sur les approches non-réductives. Et puisque pour Kim la relation de réalisation physique conduit également à l'identification en vertu de son modèle de la réduction fonctionnelle, cette relation présente le même avantage explicatif que la réduction.

Mais est-ce que Kim a raison d'affirmer que la réduction et la réalisation sont supérieures aux relations non-réductives du point de vue explicatif? On pourrait en douter en se rappelant l'argument du fossé explicatif que j'ai présenté dans la section sur les qualia du chapitre 1. En effet, l'argument de Levine s'adresse justement au réductionnisme. Levine commence par accepter la réduction, c'est-à-dire que dans le cadre de son argument du fossé explicatif il écarte les scrupules du genre "perte de quelque chose d'essentiel au mental" pour s'intéresser aux conséquences. Levine soutient que même si le réductionnisme était vrai, nous ne pourrions pas vraiment rendre compte du lien entre les qualia et leurs bases physiologiques. Par exemple, supposons que la sensation de douleur soit identifiée avec l'activation de fibres-C. Levine fait valoir qu'il reste impossible d'expliquer pourquoi la douleur consiste en la sensation particulière qui la caractérise (pourquoi est-ce qu'une sensation de brûlure est ressentit comme elle l'est plutôt que comme une sensation de picotement?), ou pourquoi cette sensation particulière est liée à cette base particulière. Le lien entre les qualia et leurs bases physiologiques serait ainsi totalement arbitraire: il s'agit d'un fait brut que nous devons accepter. Certes, l'argument de Levine repose sur nos intuitions, mais l'argument de Kim aussi. Kim ne veut pas accepter une dépendance ou une survenance comme un fait brut, mais ne voit pas de problème à accepter une identification ou une réalisation comme un fait brut.

Il me semble qu'il existe un certain flou quant à ce qui constitue une explication satisfaisante pour un phénomène donné. Cette difficulté peut se résumer par cette question: quand peut-on arrêter de demander 'pourquoi'? En effet, il semble qu'il soit toujours possible de demander 'pourquoi' peu importe la réponse qui a pu être donnée. La seule façon d'arrêter cette suite infinie de pourquoi semble ainsi d'accepter à un moment la réponse que l'on reçoit comme un fait brut. La physique contemporaine, la science qui sert souvent de modèle à toutes les autres, est remplie de faits bruts qu'il faut accepter sans autre explication. Pourquoi le champ gravitationnel décroît-il en fonction du carré de la distance plutôt que du cube? Pourquoi la vitesse de la lumière dans le vide est-elle de 300 000 km/s? Mais si l'acceptation de faits bruts est si fréquente en physique, pourquoi ne pourrait-on faire de même avec le problème corps-esprit? Kim pourrait ainsi affirmer qu'il est tout à fait naturel d'arrêter de demander 'pourquoi' à un moment donné et que l'identification de M et P constitue un tel moment. Toutefois,

l'acceptation d'un fait brut ne signifie pas que l'on néglige les incohérences ou les inconvénients qu'une position génère. Ainsi, il est toujours possible de critiquer le réductionnisme. De plus, un partisan de la survenance pourrait tout aussi bien soutenir qu'il faut accepter la survenance de *M* sur *P* comme un fait brut. La portée explicative que Kim accorde à l'identification n'est donc pas un fait objectif. Il s'agit d'une caractéristique de la pensée de Kim qui le distingue des autres philosophes et qui contribue à la bonne opinion qu'il possède de la réduction. C'est cette bonne opinion à propos de la réduction qui fait que Kim a proposé deux modèles réductifs successifs.

Retour à la réduction fonctionnelle

On a vu que le premier modèle réductif de Kim, l'identification disjonctive, présente de nombreuses difficultés et ce, même indépendamment de la question des qualia. Pour sa part, le second modèle, celui de la réduction fonctionnelle, semble moins problématique. Reste que deux questions concernant ce second modèle demandent à être examinées: la question de l'attribution de pouvoirs causaux aux états mentaux et celle de la place des qualia.

Efficacité causale des concepts de second ordre

La question de la possibilité d'accorder des pouvoirs causaux aux états mentaux à l'intérieur du modèle de la réduction fonctionnelle provient de ce que les états mentaux se révèlent n'être que des concepts de second ordre sans existence ontologique indépendante de celle de leurs réalisateurs. Il semble en effet étrange d'attribuer des pouvoirs causaux à des concepts de second ordre. Pourtant, Kim affirme que sous le modèle de la réduction fonctionnelle, les états mentaux disposent bien de pouvoirs causaux; en fait, des mêmes pouvoirs causaux que ceux de leurs réalisateurs: "...une propriété mentale est causalement efficace aussi longtemps que, et dans la mesure où, chaque réalisateur possible est causalement efficace; et une occurrence particulière d'une propriété mentale possède exactement la même efficacité causale que son

réalisateur à ce moment. Ceci confirme l'efficacité causale des propriétés mentales."[44] Ainsi, en permettant l'identification des concepts de second ordre avec des propriétés de base, le principe de l'héritage des pouvoirs causaux rendrait les états mentaux causalement efficaces.

Cependant, "avoir les mêmes pouvoirs causaux" peut être une expression ambiguë. Par exemple, si je dis que je possède la même automobile que mon conjoint, je peux vouloir dire deux choses: soit que mon conjoint et moi avons chacun une voiture du même modèle (il y a alors deux voitures dans le stationnement), soit que mon conjoint et moi avons acheté une automobile ensemble (il n'y a alors qu'une seule voiture dans le stationnement). Or, en ce qui concerne l'efficacité causale des états mentaux, la situation semble plus conforme à ce dernier cas qu'au premier. Car si les propriétés mentales ne sont que des concepts ou des appellations de second ordre, un changement dans la quantité des propriétés mentales que possède un système ou un individu n'affectera en rien sa capacité causale. Voici un exemple qui devrait illustrer ce point. Disons qu'un individu x possède la croyance que les armes à feu sont dangereuses, c'est-à-dire le concept de second ordre F. D'après le modèle de la réduction fonctionnelle de Kim cela revient à dire que x possède le réalisateur de F, soit R_F. Mais supposons que nous remplacions chez x la croyance que les armes à feu sont dangereuses par *plusieurs croyances*, soit la croyance que les revolvers sont dangereux (concept de second ordre G), la croyance que les fusils sont dangereux (concept de second ordre H), la croyance que les mousquets sont dangereux (concept de second ordre J), etc., ayant chacune leur réalisateur propre, soit respectivement R_G, R_H, R_J, etc. L'union de ces différentes croyances étant équivalent à F (la croyance que les armes à feu sont dangereuses), il y a une certaine plausibilité à supposer que *l'union* des réalisateurs R_G, R_H, R_J, etc. n'est autre que R_F, le réalisateur de F. Dit d'une autre manière, R_G, R_H, R_J, etc. sont chacun une partie de R_F. Par conséquent, x est passé d'une seule croyance, F, à plusieurs croyances, G, H, J, etc., mais pourtant, dans les faits, cela *n'a pas* entraîné de gain de pouvoirs causaux pour x, puisque les pouvoirs causaux de R_F ne sont pas différents de ceux de l'union de R_G, R_H, R_J, etc. Il est

---

[44] "*...a mental property is causally efficacious as long as, and to the extend that, each of its possible realizers is causally efficacious; and a particular instance of a mental property has exactly the causal efficacy of its realizer on that occasion. This confirms the causal efficacy of mental properties.*" Kim (1993b), p. 363.

également possible de produire un exemple de ce genre pour le cas inverse, c'est-à-dire montrant qu'une diminution du nombre de concepts de second ordre s'appliquant à un système ou un individu ne diminue pas sa capacité causale. Par conséquent, si le nombre d'états mentaux que possède un individu n'a aucun effet sur les pouvoirs causaux qu'il possède, on voit mal la pertinence de ces concepts de second ordre dans la formulation d'explications. Car en quoi est-on justifié de dire que tel individu a produit un effet physique *P en vertu du fait qu'il possède la propriété mentale (ou concept de second ordre) M?* On peut même soutenir que Kim concède ce point lorsqu'il dit que de tels concepts de second ordre ne servent que lorsque nous ne sommes pas capables ou ne voulons pas utiliser des concepts de premier ordre pour ce dont nous voulons parler (rappelons-nous l'exemple des produits chimiques différents qui sont regroupés sous le vocable de 'somnifère'). En effet, si l'emploi des concepts de second ordre ne dépend que de considérations pragmatiques, c'est qu'elles ne présentent pas d'intérêt métaphysique.

Cependant, l'argument selon lequel le nombre des états mentaux devrait avoir un impact sur les pouvoirs causaux que possède un individu est erroné. En effet, cet argument est basé sur le fait que la possession d'états mentaux réalisés par les mêmes bases physiques est exclusive. Or, rappelons que la possession d'un concept de second ordre *M* revient à la possession d'une propriété de premier ordre *P* satisfaisant une condition *C* et que cette condition spécifie les causes et effets typiques de *M*. Il est vrai que le réalisateur *P* dans son entier ne peut correspondre à plus d'un état mental, car si deux états mentaux sont réalisés par exactement le même réalisateur, ils vont avoir exactement les mêmes pouvoirs causaux et par conséquent ne constituer qu'un seul et même état mental. Mais si *P* se trouve à correspondre aux conditions de plus d'un concept de second ordre parce que des *parties* de *P* leur servent de réalisateurs , alors la possession d'états mentaux réalisés par les mêmes bases physiques *n'est pas exclusive*. C'est ce qui se passe dans l'exemple précédent: l'état mental *F* est réalisé par R_F qui n'est autre que l'union des réalisateurs R_G, R_H, R_J, etc. Dans ce cas, *F* ne possède pas exactement le même réalisateur que *G*, *H* ou *J* et ne représente donc pas le même état mental que l'un ou l'autre de ceux-ci. Mais R_F, son réalisateur, n'est pas indépendant des réalisateurs de *G*, *H*, *J*, etc. Malgré cela, la possession de *F* n'est pas exclusive à celle de *G*, *H*, *J*, etc. Une

variation dans le nombre des 'propriétés mentales' attribuées à un individu n'indique donc pas qu'il doit y avoir une variation dans le nombre de ses pouvoirs causaux. Il n'y a pas de variation dans le nombre de pouvoirs causaux parce qu'il n'y a pas véritablement de variation dans le nombre de réalisateurs: une division différente du nombre d'états mentaux et parallèlement du nombre de leurs réalisateurs n'implique pas l'apparition ou la disparition de la moindre connexion dans le cerveau du sujet. En effet, tout cet argument est basé sur une relation tout/partie. Reprendre l'argument pour montrer qu'une variation du nombre de realisateurs n'implique pas une variation du nombre de pouvoirs causaux n'aurait ainsi aucun intérêt, car cet argument implique seulement un changement conceptuel, pas ontologique. L'argument est cependant intéressant lorsqu'appliqué aux concepts de second ordre, puisqu'eux ne sont rien d'autre que des concepts. Ils sont donc affectés par cette variation de nombre. De plus, que les concepts de second ordre ne présentent pas d'intérêt métaphysique, mais seulement un intérêt pragmatique, est effectivement en accord avec la proposition de Kim. Cela revient à dire que le mental n'a plus de statut ontologique. Il n'y a pas de surprise à cela: il a déjà été dit que la réduction fonctionnelle s'opposait au dualisme des propriétés. Le mental n'est donc plus qu'un terme désignant une classe particulière de phénomènes physiques, plus particulièrement biologiques.

Qualia et fonctionnalisation

J'ai mentionné la remarque de Kim selon laquelle la réduction fonctionnelle n'est applicable que si les états mentaux - tous les états mentaux - sont fonctionnalisables. Or, si on peut raisonnablement espérer fonctionnaliser les états mentaux intentionnels, les qualia, pour leur part, apparaissent représenter une plus grande difficulté. C'est d'ailleurs là l'opinion de Kim lui-même:

> Je suis de ceux qui considère que le principal problème provient des qualia. Contrairement à ce qui en est avec les phénomènes intentionnels, nous semblons capables de concevoir sans trop de difficulté une réplique exacte de notre monde pour lequel les qualia sont distribués différemment (mondes de qualia inversés) ou complètement absents ("mondes de zombies"), bien que cette dernière possibilité soit

plus controversée. Bref, il me semble que les qualités sensibles, phénoménales des expériences, ou qualia, sont des propriétés intrinsèques si quoi que ce soit peut en être.[45]

Si les qualia sont des qualités exclusivement intrinsèques, il est alors impossible de les fonctionnaliser; car fonctionnaliser, c'est rendre extrinsèque. L'argument des qualia inversés et celui des qualia absents que j'ai présentés dans la section sur les qualia (chapitre 1) visent justement à montrer que les qualia ne produisent aucune différence fonctionnelle chez un individu ou système. Les états mentaux intentionnels, au contraire, produisent une différence puisqu'ils dirigent nos actions d'après l'opinion commune (que ce soit en philosophie ou en psychologie, populaire ou scientifique). C'est ainsi qu'alors que nous disposons de certaines pistes de recherche, particulièrement en psychologie cognitive, pour expliquer les états intentionnels, nous n'avons aucune piste pour expliquer les qualia. C'est sur cette constatation que repose la conviction de Kim que les qualia pourraient bien être intrinsèques et donc ne pas pouvoir être fonctionnalisés:

> Je trouve significative la différence intuitive suivante entre les qualia et l'intentionnalité: si quelqu'un devait nous demander de créer une chose consciente, c'est-à-dire quelque chose pouvant ressentir des douleurs, démangeaisons, chatouillements, et ainsi de suite, la seule chose que nous pourrions faire, il me semble, serait de faire une copie appropriée d'une structure - présumément un organisme biologique comme un humain ou un chat - dont nous savons, ou croyons, qu'elle est consciente et capable de ressentir ces sensations. Nous sommes incapables de concevoir, à partir d'un raisonnement théorique, une sorte complètement nouvelle de structure dont nous pourrions prévoir qu'elle sera

---

[45] "*I am with those who believe that the main trouble comes from qualia. Unlike the case of intentional phenomena, we seem able, without much difficulty, to conceive an exact physical duplicate of this world in which qualia are distributed differently (worlds with qualia inversions) or entirely absent ("zombie worlds"), although the latter possibility is more controversial. To get to the point without fuss, it seems to me that the felt, phenomenal qualities of experiences, or qualia, are intrinsic properties if anything is.*" Kim (1998), p. 101-102.

consciente; je ne pense pas que nous sachions même par où commencer, ou encore comment juger de notre succès.[46]

Il est vrai que les qualia apparaissent posséder un aspect intrinsèque, mais le physicalisme non-réductif considère également que les qualia exercent des pouvoirs causaux. Or, puisque fonctionnaliser revient essentiellement à définir quelque chose par ses pouvoirs causaux, on pourrait établir le critère suivant:

> **Critère de la fonctionnalisation:** Toute chose (état, événement, etc.) possédant des pouvoirs causaux est fonctionnalisable, c'est-à-dire qu'elle peut être définie en termes de fonction.

Selon le critère de la fonctionnalisation, si les qualia possèdent des pouvoirs causaux, alors ils sont fonctionnalisables. Et s'ils n'en possèdent pas, alors ils ne sont que des épiphénomènes que nous pouvons aussi bien éliminer. Bien sûr, définir quelque chose par ses pouvoirs causaux, c'est négliger toutes ses autres caractéristiques: fonctionnaliser, c'est effectuer une sorte de simplification ou de réduction, même s'il ne s'agit pas d'une réduction à quelque chose de physique. Lorsque l'on fonctionnalise un état mental intentionnel, tout ce que le fonctionnalisme retient de cet état, ce sont ses relations avec les autres états mentaux ou physiques. Évidemment, le fonctionnalisme soutient que les états mentaux intentionels ne sont *rien d'autres* que des états fonctionnels; ou du moins, plus modestement, que rien d'important n'est perdu lors de la fonctionnalisation des états mentaux intentionnels. Le problème avec les qualia est celui-ci: c'est justement la partie essentielle des qualia qui est éliminée par leur fonctionnalisation. En fait, tout pouvoir causal que pourrait éventuellement posséder un quale peut en être séparé conceptuellement. Par exemple, supposons que la douleur (état mental *M*) soit typiquement produite par une lésion à une partie du corps (état physique *P*) et qu'elle cause

---

[46] "*I find the following intuitive difference between qualia and intentionality significant: If someone should ask us to create a device with consciousness, say something that can fell pain, itch, tickle, and the like, the only thing we can do, it seems to me, is to make an appropriate duplicate of a structure, presumably a biological organism like a human or a cat, that we know, or believe, to be conscious and capable of these sensations. We are not capable of designing, through theoretical reasoning, a wholly new kind of structure that we can predict will be conscious; I don't think we even know how to begin, or indeed how to measure our success.*" Kim (1998), p. 102.

typiquement des réactions telles que des cris et des lamentations (état physique *Q*) de même qu'une tendance à éviter la source de la lésion (état physique *R*). Il est alors facile d'obtenir une description fonctionnelle de la douleur à partir d'une telle caractérisation: un état *M* est un état de douleur si *P* cause *M* et que *M* cause *Q* et *R*. Seulement une telle description laisse de côté tout ce que la douleur a de phénoménal. En effet, par définition, *M* devient l'état causé par *P* et causant *Q* et *R*: aucune exigence sur la nature intrinsèque de *M* (comme le fait d'être un état phénoménal de douleur) n'est imposée. Ainsi, il nous est possible de créer un système respectant la description fonctionnelle de *M*, c'est-à-dire où *M* est causée par la présence de lésions et où *M* entraîne la présence de cris et de lamentations de même qu'une tendance à éviter la source des lésions, sans pourtant que nous ne soyons portés à considérer que le système en question (car il peut s'agir d'un montage électronique relativement simple) présente la moindre conscience phénoménale. Les partisans des qualia demandent alors: en quoi la douleur est-elle de la douleur si elle ne fait pas *mal* (au sens phénoménal du terme)?

Il peut être tentant de simplement ajouter l'aspect phénoménal au modèle fonctionnaliste, mais il ne semble pas y avoir de façon cohérente de le faire. Pour continuer avec l'exemple de la douleur, si nous posons que *M* lui-même doit présenter les qualités phénoménales intrinsèques de la douleur, nous ne pouvons alors dire que nous avons obtenu une fonctionnalisation de la douleur et prétendre aux avantages conceptuels du fonctionnalisme. Et si nous supposons que le quale de douleur est un effet de *M* (comme semble le faire Searle en disant que la conscience phénoménale est causée par le cerveau[47]), alors nous nous retrouvons avec un tas de problèmes conceptuels concernant ce nouvel état qui, dans le meilleur des cas, se révélera un épiphénomène.

La concience phénoménale représente ainsi une grande difficulté pour le fonctionnalisme. Ce qui, d'une certaine façon, est ironique: le fonctionnalisme réussit à rendre compte sans trop de difficultés des états mentaux intentionnels qui sont à la base de la rationalité et donc propres à des organismes hautement évolués, mais il échoue à expliquer la concience phénoménale, un phénomène qui se trouve probablement présent chez des êtres relativement simples. Cet échec

[47] Searle (1992).

du fonctionnalisme se répercute tel quel dans le modèle de la réduction fonctionnelle de Kim: s'il est impossible de fonctionnaliser les qualia, il est impossible de les réduire. Kim reconnaît cette difficulté: "la mauvaise nouvelle est que certaines propriétés mentales, particulièrement les propriétés phénoménales des expériences conscientes, semblent résister à la fonctionnalisation et cela signifie qu'il n'y a aucun moyen de rendre compte de leur efficacité causale à l'intérieur d'une optique physicaliste."[48] Est-ce à dire qu'il nous faut choisir entre (1) adopter le modèle de la réduction fonctionnelle pour *tous* les états mentaux même si cela implique éliminer les qualia ou (2) abandonner le physicalisme? Peut-être pas. En effet, dans *Mind in a Physical World*, Kim laisse entendre que *l'émergentisme* pourrait bien constituer la solution au problème des qualia: "il me semble que si l'émergentisme est approprié pour quoi que ce soit, il est plus probable que ce soit à propos des qualia qu'à propos de quoi que ce soit d'autre."[49] Un examen de cette possibilité s'impose donc avant de conclure notre recherche.

L'émergentisme

L'émergentisme ne constitue pas un courant dominant de la philosophie de l'esprit contemporaine, c'est pourquoi je ne l'ai pas présenté au chapitre 1. Il s'agit d'un courant dont les origines remontent au milieu du 19e siècle et dont l'âge d'or se situe au début du 20 e siècle. Malgré tout, l'émergentisme semble s'incrire parfaitement dans le courant du physicalisme non-réductif contemporain, ce qui explique sans aucun doute le regain d'intérêt dont il est l'objet de la part des philosophes depuis quelques années[50], un mouvement probablement en expansion pour les prochaines années. Cet accord entre l'émergentisme et le physicalisme non-réductif pourrait cependant n'être qu'apparent. En effet, deux interprétations divergentes de

---

[48] "*the real bad news is that some mental properties, notably phenomenal properties of conscious experiences, seem to resist functionalization, and this means that there is no way to account for their causal efficacy within a physicalist scheme.*" Kim (1998), p. 118-119.

[49] "*it seems to me that if emergentism is correct about anything, it is more likely to be correct about qualia than about anything else.*" Kim (1998), p. 103.

[50] *En témoigne notamment la parution du recueil* Emergence or Reduction?, A. Beckermann, H. Flohr et J. Kim, dirs., (Berlin: De Gruyter, 1993) et d'un nombre croissant d'articles sur l'émergentisme. Un défenseur important de l'émergentisme moderne dans le cas du problème corps-esprit est Paul Humphreys (1996, 1997a, 1997b).

l'émergentisme peuvent être adoptées quant à savoir si cette approche respecte ou non la fermeture du monde physique. Si l'émergentisme respecte la fermeture du monde physique, alors cette approche se situe bien dans le courant du physicalisme non-réductif. Mais si au contraire l'émergentisme ne respecte pas la fermeture du monde physique, alors il s'agit d'une thèse plus radicalement différente qui ne fait pas partie du consensus de physicalisme non-réductif. Je reviendrai plus loin sur cette question. À mon avis, toutefois, Kim considère que l'émergentisme respecte la fermeture du monde physique, car autrement il n'accepterait pas d'adopter une telle approche compte tenu de sa grande conviction dans le physicalisme.

Est-ce que l'émergentisme constitue une approche intéressante pour le cas des qualia, comme semble le dire Kim? Dans un article en voie de publication, Kim affirme de façon plus claire son opinion sur le sujet:

> Il me semble que si quoi que ce soit est émergent, les propriétés phénoménales de la conscience, ou "qualia", constituent les canditats les plus prometteurs. Je ne veux pas ici reprendre les arguments habituels pour et contre, mais simplement affirmer, pour ce qu'il vaut, mon propre biais envers la classe des 'pour': les qualia sont des propriétés intrinsèques si quoi que ce soit en est, et les fonctionnaliser revient à les éliminer en tant que propriétés intrinsèques.[51]

Pour que l'émergentisme soit d'une quelconque utilité, cependant, il faut qu'il représente une approche qui puisse rendre compte des qualia d'une manière qui, d'une part, préserve le caractère intrinsèque des qualia et qui, d'autre part, confère des pouvoirs causaux aux qualia. À première vue, ces conditions sont compatibles avec l'idée générale de l'émergentisme selon laquelle les structures se sont complexifiées au fil de l'évolution et qu'à mesure que furent atteints des niveaux supérieurs d'organisation, de *nouvelles propriétés* apparaissaient. D'après les émergentistes, ces nouvelles propriétés transcendent, d'une certaine façon, les propriétés des parties constituantes des systèmes. Les systèmes complexes réagissent alors d'une manière qui n'aurait pu être prédite sur la base des lois dirigeant les systèmes plus simples. Je vais donc

---

[51] "*It seems to me that if anything is going to be emergent, the phenomenal properties of consciousness, or "qualia", are the most promising candidates. Here I don't want to rehearse the standard arguments pro and con, but merely affirm, for what it's worth, my own bias toward the pro side: qualia are intrinsic properties if anything is, and to functionalize them is to eliminate them as intrinsic properties.*" Kim, manuscrit, à paraître dans Philosophical Studies, p. 17.

présenter les idées principales de l'émergentisme, communes tant à l'émergentisme classique qu'à l'émergentisme moderne, et tenter de voir si une véritable nouvelle voie est tracée par cette approche.

Les cinq thèses principales de l'émergentisme

1- L'émergence des entités complexes

L'émergentisme s'intéresse aux propriétés des entités complexes. Pour cela, il faut évidemment poser que de telles entités existent et dire d'où elles proviennent. C'est ce que la première thèse de l'émergentisme nous dit:

> **Émergence des entités complexes de niveau supérieur:** Des systèmes ayant un niveau supérieur de complexité émergent de la combinaison d'entités de niveau inférieur de complexité en une configuration structurelle nouvelle.[52]

Cette thèse de l'émergence des entités complexes n'est pas exclusive à l'émergentisme. En fait, puisqu'il n'est pas fait mention de propriétés émergentes dans cette thèse, l'emploi du mot 'émergence' est plus heuristique qu'autre chose. Un non-émergentiste peut reprendre cette thèse telle quelle en changeant simplement le mot 'émergence' par 'apparition'.

2- L'émergence des propriétés de haut niveau

Chez ces entités complexes, l'émergentiste trouve deux types de nouvelles propriétés, soit les propriétés émergentes et les propriétés résultantes, alors qu'un non-émergentiste ne reconnaîtrait l'existence que des seules propriétés résultantes:

> **Émergence des propriétés de haut niveau:** Toutes les propriétés des entités de niveau supérieur trouvent leur origine dans les propriétés et relations qui caractérisent leurs

---

[52] "1. *Emergence of complex higher-level entities*: *Systems with a higher-level of complexity emerge from the coming together of lower-level entities in new structural configurations.*" Kim, manuscrit, à paraître dans Philosophical Studies, p. 20.

parties constituantes. Certaines propriétés de ces systèmes complexes sont "émergentes" alors que les autres ne sont que "résultantes.[53]

D'après la thèse de l'émergence des propriétés de niveau supérieur, tant les propriétés émergentes que les propriétés résultantes sont dépendantes de leurs 'propriétés de base', c'est-à-dire des propriétés possédées par les parties constituantes de l'entité complexe à laquelle elles appartiennent. Si les qualia sont des propriétés émergentes, leurs propriétés de bases seront évidemment des propriétés physiques. Cette relation de dépendance entre propriétés émergentes et résultantes d'un côté et conditions de base de l'autre implique au minimum la survenance faible, sinon la forte. En effet, pour être d'une quelconque utilité, la notion d'émergence doit impliquer une relation de dépendance qui vaut au moins pour un monde. Il s'agit cependant là d'un point commun entre les propriétés émergentes et les propriétés résultantes: jusqu'à maintenant, nous n'avons rien appris sur la différence existant entre les propriétés émergentes et résultantes. Il est maintenant temps d'y remédier: les trois dernières thèses de l'émergentisme concernent les principales caractéristiques des propriétés émergentes.

3- L'imprévisibilité des propriétés émergentes

La première caractéristique unique des propriétés émergentes, et la troisième thèse de l'émergentisme, est que l'on ne peut prédire la présence des propriétés émergentes à partir de la connaissance des propriétés de base:

> **Imprévisibilité des propriétés émergentes:** Les propriétés émergentes ne sont pas prévisibles à partir d'une connaissance exhaustive de leurs 'conditions de base'. Inversement, les propriétés résultantes sont prévisibles à partir d'une telle connaissance du niveau inférieur.[54]

---

[53] "2. *Emergence of higher-level properties*: *All properties of higher-level entities arise out of the properties and relations that caracterize their constituent parts. Some properties of these higher, complex systems are "emergent", and the rest merely "resultant".*" Kim, manuscrit, à paraître dans Philosophical Studies, p. 20.

[54] "3. *The unpredictability of emergent properties*: *Emergent properties are not predictable from exhaustive information concerning their "basal conditions". In contrast, resultant properties are predictable from lower-level information.*" Kim, manuscrit, à paraître dans Philosophical Studies, p. 20.

Prenons un exemple: un individu x possède les propriétés Q et R, Q étant la propriété *émergente* d'avoir une sensation de douleur et R étant la propriété *résultante* d'avoir un poids de 65 Kg. Bien que R constitue une propriété de x qu'aucune partie de x ne possède, la présence de R chez x peut être prédite sur la base de la connaissance du poids de toutes les parties de x, disons le poids de chacune de ses cellules. En effet, il suffit alors d'additionner le poids de toutes les cellules de x pour obtenir 65 Kg., c'est-à-dire R (pour plus de simplicité, je néglige le poids de certains composants de R qui ne sont pas constitués de cellules, comme le dernier repas que x est en train de digérer). Une propriété résultante n'est pas nécessairement *facile* à prédire, mais elle est *en principe* prévisible. Par contre, Q, en tant que propriété émergente, n'est pas prévisible sur la base d'une connaissance complète de la base physique sur laquelle survient la douleur, disons l'activation de fibres-C. Le cas est similaire à celui de l'exemple de Jackson présenté dans la section sur les qualia (chapitre 1): Mary, la spécialiste de la vision, qui sait tout ce qu'il y a à savoir de physique (c'est-à-dire tout sur les conditions de base) à propos de la vision des couleurs, mais qui a toujours vécu dans un environnement monochrome, ne sait pas 'ce que cela fait' de voir du rouge. Certes, Mary pourrait prévoir qu'elle va avoir une impression de rouge en regardant une tomate mûre, mais elle ne peut prévoir "ce que cela fait" que d'avoir un quale de rouge. De façon similaire, d'après la thèse de l'imprévisibilité des propriétés émergentes, un spécialiste du système nerveux pourrait savoir tout ce qu'il y a savoir de physique sur les fibres-C sans pour autant prévoir que l'activation d'une fibre-C conduit à l'apparition d'un quale de douleur. Remarquez que j'ai parlé d'un spécialiste du système nerveux et non pas d'un spécialiste de la douleur, car un spécialiste de la douleur saurait, d'une certaine manière, que l'activation de fibres-C conduit à l'apparition d'un quale de douleur, tout comme Mary peut prévoir la présence d'un quale de rouge lorsqu'un sujet observe une tomate mûre. C'est qu'il faut distinguer entre une *prédiction inductive* et une *prédiction théorique*. Une prédiction inductive n'implique pas de connaissances approfondies du lien pouvant exister entre deux phénomènes: on peut prédire la pluie à partir de la vision de gros nuages gris, sans savoir que les nuages sont constitués de vapeur d'eau. Lorsque Mary prédit qu'un sujet aura une impression de rouge suite à la vision d'une tomate mûre, c'est une prédiction inductive qu'elle fait. En effet, elle sait qu'il y a une

corrélation entre l'activation d'une structure donnée du cerveau, disons *S*, et la vision du rouge, mais elle ne peut pas *expliquer* cette prédiction en référant à la corrélation entre la vision du rouge et *S*, car il s'agit d'un fait brut. Il a été observé que la pluie a toujours lieu lorsque sont présents de gros nuages gris et que la perception d'un quale de rouge a toujours lieu lorsque *S* est activé.

De même, les émergentistes ne nient pas qu'il soit possible de prédire *inductivement* l'apparition d'un quale de douleur sur la base de l'activation de fibres-C (ou du fait que *x* s'est brûlé le doigt, sachant qu'une brûlure conduit à l'activation de fibres-C). De fait, il est possible d'établir toute une série de lois inductives de ce genre. Ce que soutiennent les émergentistes, c'est qu'il est impossible de prévoir *théoriquement* la présence d'une propriété émergente, c'est-à-dire à partir d'une connaissance exhaustive des conditions de base, par exemple des fibres-C, mais *sans la possesion d'une loi inductive* du genre décrit précédemment. On effectue une prédiction théorique lorsque l'on prédit un phénomène sur la base de notre théorie. Par exemple, la théorie de la relativité générale d'Einstein prédisait que la lumière pouvait être déviée par une masse malgré que la lumière elle-même n'ait pas de masse et ce, en opposition avec la loi de la gravitation de Newton. Aucune observation de ce phénomène n'avait été faite au moment où Einstein a élaboré sa théorie, aussi il ne s'agissait pas d'une prédiction inductive, mais bien théorique.

Bref, il ne fait aucun doute que nous pouvons prévoir inductivement l'apparition de qualia, mais nous sommes apparemment incapables de les prévoir théoriquement, c'est-à-dire sur la base de la connaissance de leurs conditions de base (une connaissance complète de la neurophysiologie du cerveau). Ainsi, les qualia ne seraient pas des propriétés résultantes, mais bien émergentes.

4- L'inexplicabilité/irréductibilité des propriétés émergentes

La quatrième thèse de l'émergentisme est celle de l'inexplicabilité et de l'irréductibilité des propriétés émergentes:

Inexplicabilité/irréductibilité des propriétés émergentes: Les propriétés émergentes, contrairement aux propriétés qui ne sont que résultantes, ne sont ni explicables par leurs conditions de base, ni réductibles à elles.[55]

Cette thèse est liée à celle de l'imprévisibilité des propriétés émergentes. En effet, si les propriétés émergentes pouvaient être expliquées ou réduites à leurs conditions de base, alors il serait possible de prédire théoriquement leur présence. Ainsi, bien que d'après la seconde thèse de l'émergentisme (la thèse de l'émergence des propriétés de niveau supérieur) les propriétés émergentes proviennent de leurs conditions de base, elles ne peuvent pas, malgré tout, être réduites à ces conditions de base, ni même expliquées par elles. L'irréducibilité des propriétés émergentes correspond bien à la thèse de la survenance et à l'esprit du physicalisme non-réductif. Toutefois, refuser également toute explication des propriétés émergentes est peut-être beaucoup demander à la plupart des philosophes. En effet, cela revient à accepter l'émergence de certaines propriétés comme un fait brut qui ne peut recevoir d'explication. C'est d'ailleurs ainsi que le concevaient les anciens émergentistes, comme Samuel Alexander qui demandait que l'on accepte l'émergence des propriétés mentales avec une "piété naturelle".

5- L'efficacité causale des propriétés émergentes

La dernière thèse principale de l'émergentisme est celle de l'efficacité causale des propriétés émergentes. En effet, dans l'hypothèse où de telles propriétés existent, il est naturel de leur attribuer des pouvoirs causaux; autrement, les propriétés émergentes ne seraient que des épiphénomènes.

---

[55] "4. *The unexplainability/irreducibility of emergent properties*: *Emergent properties, unlike those that are merely resultant, are neither explainable nor reducible in terms of their basal conditions.*" Kim, manuscrit, à paraître dans Philosophical Studies, p. 21.

Efficacité causale des propriétés émergentes: Les propriétés émergentes possèdent leurs propres pouvoirs causaux, des pouvoirs causaux nouveaux et irréductibles à ceux de leurs constituants de base.[56]

C'est l'attribution de pouvoirs causaux irréductibles à ceux de leurs bases qui fait des propriétés émergentes de véritables ajouts à l'ontologie. Ainsi, les émergentistes refuseraient la fonctionnalisation des propriétés émergentes: celles-ci ne sont pas que des concepts de second ordre, mais bien des propriétés intrinsèques. S'ils sont des propriétés émergentes, alors les qualia pourraient échapper à la fonctionnalisation tout en possèdant bien des pouvoirs causaux.

Cependant, si les propriétés émergentes possèdent des pouvoirs causaux, alors il semble qu'elles peuvent être fonctionnalisées en vertu du *critère de la fonctionnalisation* selon lequel toute chose (état, événement, etc.) possédant des pouvoirs causaux est fonctionnalisable. La même discussion présentée dans la section sur les qualia et la fonctionnalisation s'applique ici: les pouvoirs causaux des qualia peuvent être fonctionnalisés, mais cela laisse inévitablement de côté les qualia eux-mêmes, c'est-à-dire que tout pouvoir causal atribué à un quale peut en être séparé conceptuellement et fonctionnalisé indépendamment. Les qualia ne peuvent être fonctionnalisés parce qu'ils sont essentiellement intrinsèques et que fonctionnaliser, c'est rendre extrinsèque.

Il ressort de cette présentation des cinq thèses principales de l'émergentisme que cette approche se situe très près du physicalisme non-réductif, du moins si on considère que l'émergentisme respecte la fermeture du monde physique. L'émergence semble également très près de la survenance, bien qu'il soit plus difficile de les comparer. En effet, nous avons des définitions relativement précises des différents types de survenance, mais leurs conséquences sont encore largement débattues. Inversement, les cinq thèses de l'émergentisme que je viens de présenter sont explicites quant aux conséquences de la relation d'émergence. Les principales

---

[56] "5. *The causal efficacy of the emergents: Emergent properties have causal powers of their own — novel causal powers irreducible to the causal powers of their basal constituents.*" Kim, manuscrit, à paraître dans Philosophical Studies, p. 21.

caractéristiques de la propriété survenante se retrouvent chez la propriété émergente, à savoir une certaine forme de dépendance au physique tout en étant irréductible à celui-ci. D'autres caractéristiques de l'émergence, l'imprédictibilité et l'inexplicabilité, ne se retrouvent pas chez la survenance, mais ne semblent pas non plus incompatibles avec elle. Enfin, la possesion de pouvoirs causaux ne fait pas explicitement partie de la définition de la survenance, mais il est clair que les propriétés survenantes sont conçues (ou du moins l'étaient) comme ayant des pouvoirs causaux, car autrement l'argument de la survenance de Kim n'aurait pas été si dommageable. Il semble donc qu'une propriété survenante ne soit pas nécessairement une propriété émergente, mais qu'une propriété émergente puisse être considérée comme une propriété survenante (encore une fois, à la condition que l'émergence respecte la fermeture du monde physique). Il est maintenant temps de voir si l'émergentisme, une fois appliqué au problème de la causalité mentale, offre réellement une alternative intéressante aux approches du physicalisme non-réductif.

L'émergentisme et le problème de la causalité mentale

J'ai mentionné dès l'introduction que le problème de la causalité mentale n'est pas celui du 'si', mais celui du 'comment'. La cinquième thèse de l'émergentisme accorde des pouvoirs causaux aux états mentaux, répondant ainsi affirmativement à la question du 'si', mais si l'émergence des propriétés mentales doit être acceptée comme un fait brut, il est difficile de voir en quoi le recours à l'émergence répond à la question du 'comment'. En effet, l'émergentisme n'offre pas de réponse à la question: "Pourquoi *A* émerge-t-il de la base *B*?". L'émergentisme se retrouve alors un peu dans la même situation que le réductionnisme: tout deux nous offrent un fait brut qu'il faut tout simplement accepter en l'absence de réponse à nos questions sur le 'pourquoi' de cette relation d'émergence ou d'identité. Je parle bien sûr ici d'identité contingente, pas d'identité nécessaire. La relation de survenance non plus n'explique pas pourquoi elle vaut entre deux phénomènes ou propriétés, mais pour Kim une réponse est supposée exister, éventuellement en termes d'une autre relation. L'émergence et l'identité diffèrent de la survenance en ce que, pour ces deux types de relation, on considère qu'aucune réponse ne peut et ne doit être fournie. L'inexplicabilité est l'une des thèses majeures de

l'émergentisme, alors que l'identité est souvent perçue comme rendant inutile toute question supplémentaire:

> Référer à des identités [...] devrait être vu [...] non pas comme répondant à ces questions explicatives, mais plutôt comme les neutralisant ou les dissipant - c'est-à-dire montrant *qu'il n'y a rien à expliquer ici.* Pourquoi est-ce qu'il y a de l'eau seulement où et quand il y a de l'H_2O? Pourquoi est-ce que les deux corrèlent? Eh bien, l'eau n'est rien d'autre que de l'H_2O et il n'y a aucune corrélation à expliquer.[57]

Le recours à l'identité dans le cas des qualia pour éviter le fossé explicatif (en fait, pour nier son existence dès le départ) est une tactique connue mais qui laisse plusieurs philosophes insatisfaits. En effet, des questions comme "pourquoi y a t-il des qualia?", "pourquoi tel quale est-il identique avec telle base neurologique plutôt qu'avec telle autre?" apparaissent comme des questions légitimes qui ne devraient pas être écartées aussi facilement. Que l'émergentisme se retrouve dans la même situation est toutefois plus surprenant, puisque cette approche se situe plus près du physicalisme non-réductif que le réductionnisme. Mais si ces demandes d'explication sont légitimes quant à l'identité, elles devraient également l'être pour l'émergence. La thèse de l'imprévisibilité des propriétés émergentes ne peut certainement pas être invoquées comme solution au problème, puisqu'elle en constitue en fait la *cause*. En effet, cette thèse proclame que les propriétés émergentes sont imprévisibles, mais ne justifie pas cette affirmation.

Au plan explicatif, le modèle de la réduction fonctionnelle de Kim est de prime abord dans une position légèrement meilleure que la simple affirmation d'identité ou d'émergence puisqu'il offre certaines réponses. Ainsi, à la question "pourquoi la propriété mentale M est-elle présente chaque fois que P est présent chez un système s?" il est possible de répondre que M est présent chaque fois que P est réalisé par un système s parce qu'avoir M c'est, par définition, avoir une propriété ayant un rôle causal C et que, dans les systèmes comme s, P est la propriété (ou l'une

---

[57] "*Appealing to identities [...] should be seen [...] not as answering these explanatory questions, but rather as neutralizing or dissipating them - that is, showing that* there is nothing here to be explained. *Why is there water just where and when H_2O is present? Why do the two correlate? Well, water just is H_2O, and there is no correlation to be explained.*" Kim, manuscrit, à paraître dans Philosophical Studies, p. 16-17. Kim souligne.

des propriétés) satisfaisant la condition *C*. Ainsi, pour les systèmes comme *s*, avoir *M* c'est avoir *P*. Ce genre de réponse peut être jugé plus explicatif que simplement dire que *M* est identique à *P* ou que *M* émerge de *P*. Mais le problème n'est que reporté d'une étape, puisque l'on pourrait demander pourquoi *M* a-t-il une propriété satisfaisant la condition *C* plutôt qu'une autre? C'est une question sans doute idiote puisque la réponse évidente est que *M est avoir une propriété satisfaisant C*. Mais c'est le même genre de réponse que produit un partisan de l'identité lorsqu'il répond que *M est P*. Si la question du pourquoi est valable dans un cas, ne devrait-elle pas l'être dans l'autre? Il y a là un important débat à faire sur nos intuitions, débat qui revient en fin de compte à préciser ce qu'est une explication et à dire quand il est légitime d'arrêter de demander des justifications pour nos réponses. De toute façon, peu importe les mérites réels ou souhaités du modèle de la réduction fonctionnelle, nous avons vu qu'il ne s'applique pas aux qualia. Qu'a donc à nous offrir l'émergentisme?

Nous avons vu que l'émergentisme présente le même défaut que le réductionnisme quant à la demande d'explication. Aussi, il ne peut s'agir là d'un avantage pour l'émergentisme. Par contre, le fait que l'émergentisme soit anti-réductionniste le rend plus attrayant pour les partisans du physicalisme non-réductif. De plus, l'émergentisme est consistant avec la thèse de la réalisation multiple, puisqu'une même propriété peut émerger de différentes bases. Enfin, la cinquième thèse de l'émergentisme accorde des pouvoirs causaux aux instances de propriétés émergentes, pouvoirs causaux qui sont indépendants de ceux que peut avoir la base des propriétés émergentes, ce qui permet à l'émergentisme de résister également à l'argument de la survenance.

Mais l'attribution de tels pouvoirs causaux aux instances de propriétés émergentes pose la question du respect de la fermeture du monde physique par l'émergentisme. En effet, est-il concevable que des propriétés qui dépendent de la présence d'une base pour exister (seconde thèse de l'émergentisme) puissent posséder des pouvoirs causaux indépendants de ces mêmes bases? Commençons par voir ce que la possession de pouvoirs causaux par une instance de propriété émergente signifie. Si une propriété émergente *E* possède des pouvoirs causaux, alors elle devrait être en mesure de causer l'apparition d'une autre propriété émergente, *E**, de même

niveau. Mais la seconde thèse de l'émergentisme nous dit que E^* possède une base, B^*. La question se pose alors: est-ce que E^* est présent parce que E a causé E^* ou est-ce que E^* est présent parce que sa base, B^*, est présente. Un problème similaire s'est présenté avec la survenance. La question avait alors été tranchée par le recours au *principe de la primauté de la base* selon lequel pour causer l'instantiation d'une propriété survenante, il faut causer l'instantiation de sa (ou d'une de ses) propriété(s) de base. Est-ce qu'un principe similaire pourrait s'appliquer ici? Un tel principe serait en accord avec la seconde thèse de l'émergentisme, mais contredirait la cinquième thèse de l'émergentisme (et peut-être également la quatrième).

D'un autre côté, même si on dit que E cause la présence de E^*, il semble inévitable que B^* doive également être présent. Cela revient à dire que E doit causer B^* en plus de E^*. Autrement dit, une propriété émergente doit non seulement pouvoir causer l'apparition de propriétés du même niveau qu'elle, mais également du niveau inférieur. Mais si E cause (apparemment) à la fois E^* et B^*, peut-être qu'en fait E ne cause que B^* et que E^* simplement émerge de B^* par la suite. Cette hypothèse présente l'avantage d'être compatible à la fois avec la seconde et avec la cinquième thèse de l'émergentisme. Cette procédure n'apparaît cependant pas très plausible. En fait, la question se pose de savoir si c'est bien E qui cause B^* en vertu des pouvoirs causaux qui lui sont attribués de façon inexplicable, ou si c'est B, la base de E, qui cause B^* de la manière familière dont un phénomène physique en cause un autre. Mais dire que B cause B^* c'est nier à E le pouvoir de causer E^*, ce qui contrevient à la cinquième thèse de l'émergentisme. Il faut donc conclure que l'émergentisme résiste à l'argument de la survenance grâce à sa cinquième thèse, qui assure que les propriétés émergentes possèdent des pouvoirs causaux différents de ceux de leur base. On n'aurait pu penser qu'une affirmation sans équivoque comme celle-ci allait simplifier la compréhension des implications de la notion d'émergence. Ce n'est malheureusement pas le cas, comme notre discussion vient de le montrer. On ne peut pas non plus soutenir qu'il faudrait éliminer l'une ou l'autre des thèses de l'émergentisme, car cela ferait perdre aux propriétés émergentes ce qui les distinguent des autres sortes de propriétés. Reste nos questions du début: est-il concevable que des propriétés qui dépendent de la présence d'une base pour exister puissent posséder des pouvoirs causaux

indépendants de ces mêmes bases? L'existence de tels pouvoirs causaux impliquent-ils un rejet du principe de la fermeture du monde physique?

Comme on l'a vu, il est impossible de répondre à ces questions sur la seule base des cinq thèses de l'émergentisme. Au moins un modèle d'émergence aurait toutefois été élaboré qui permettrait apparemment à l'émergentisme de respecter la fermeture du monde physique et qui ferait sens de l'attributon de pouvoirs causaux aux propriétés émergentes qui soient indépendants de ceux de leur base. Ce modèle a été présenté par Paul Humpreys dans une série d'articles, dont en particulier "How Properties Emerge" (1997b). L'idée centrale de Humpreys, c'est de voir l'émergence comme un processus plutôt que comme un état. Il soutient qu'il est vrai que pour apparaître une propriété émergente nécessite une base, mais qu'à partir du moment où la nouvelle propriété émerge, la base n'a plus d'existence indépendamment de la nouvelle propriété. Humpreys parle alors d'un processus de 'fusion'. Plus précisément, pour qu'une nouvelle propriété émergente de niveau i+1 apparaisse, il faut que *deux* instances de niveau i fusionnent ensemble. Ni l'une, ni l'autre n'a alors d'existence indépendante. Cette nouvelle propriété ainsi créée est alors identique avec la propriété émergente de niveau i+1. Ce processus de fusion est supposé réellement se produire dans le monde: "*Lorsque je parle de fusion, je parle d'une opération physique réelle et non pas d'une opération logique ou mathématique sur des représentations prédicatives de propriétés.*"[58] Puisque la (double) base et la nouvelle propriété ne font qu'une, les problèmes soulignés plus haut disparaissent: il n'y a pas de compétition entre "E cause E*" et "B, la base de E, cause B*, d'où émerge E*" si E et B, tout comme E* et B*, sont identiques.

La première chose à remarquer concernant ce modèle, cependant, est qu'il en est encore au stade de l'ébauche. Beaucoup de travail reste encore à faire. Deuxièmement, Humpreys se concentre pour l'instant sur les propriétés émergentes non-mentales, c'est-à-dire certaines propriétés physiques, chimiques ou biologiques. Reste ne prouve donc que la proposition de Humpreys sera applicable aux états mentaux en général et aux qualia en particulier.

---

[58] "*By a fusion operation, I mean a real physical operation, and not a mathematical or logical operation on predicative representations of properties.*" Humpreys (1997b), p.10. En italiques dans le texte.

Troisièmement, Humpreys précise que, selon ce modèle de l'émergence (les propriétés de niveau *i* étant les propriétés physiques du niveau le plus élémentaire): "Nous perdons, bien sûr, la fermeture causale des instances de propriétés de niveau *i*."[59] . Malgré cela, la proposition de Humphreys ne revient pas automatiquement à briser la fermeture causale du monde physique, à moins d'avoir une conception très étroite du physique. En effet, le niveau $i + 1$ est fort probablement encore quelque chose de physique selon l'appréciation commune.

Mais si le modèle de Humpreys implique la perte de la fermeture causale du niveau *i*, ne peut-on pas s'attendre à ce que pour n'importe quel niveau donné $i + n$, il soit possible de fusionner deux instances de ce niveau pour créer une nouvelle propriété émergente qui briserait la fermeture causale de ce niveau? Si cela est vrai, alors deux options s'offrent à nous: soit l'émergence des propriétés mentales est possible, soit elle ne l'est pas. Si elle est possible, nous perdons la fermeture causale du monde physique. Si elle ne l'est pas, nous perdons le mental d'une manière similaire à celle de l'éliminativisme ou de la réduction. Dans ce dernier cas, nous ne pouvons pas rendre compte de la causalité mentale grâce à l'émergentisme. De plus, si l'émergentisme est considéré comme contrevenant à la thèse de la fermeture causale du monde physique, le choix de l'émergentisme pour expliquer les qualia serait assez étonnant dans le cas de Kim. En effet, pourquoi s'être donné tant de mal à élaborer un modèle respectant la fermeture du monde physique dans le cas des états mentaux intentionnels (celui de la réduction fonctionnelle), si c'est pour ensuite rejeter la fermeture du monde physique pour les états mentaux phénoménaux? Ce serait donc dire que la seule piste de recherche laissée ouverte par Kim pour rendre compte des qualia, l'émergentisme, remettrait en cause la stabilité de la position de Kim quant aux états mentaux intentionnels. Ce serait là changer un mal pour un autre pire encore.

Peut-être que le modèle de Humpreys, ou bien peut-être un autre modèle d'émergence, réussira à éviter ces inconvénients. Si c'était le cas, Kim pourrait alors recourir à l'émergentisme pour rendre compte des qualia. Quel serait cependant l'utilité du modèle de la réduction fonctionnelle des états mentaux intentionnels si un modèle non-réductif était disponible? Ne

[59] "*We do, of course, lose the causal closure of i-level property instances.*" Humphreys (1997b), p. 14.

peut-on pas penser qu'il serait plus simple de considérer tous les états mentaux comme émergents?

Conclusion

Notre conviction en la possibilité de la causalité mentale nous pousse à trouver comment elle peut être possible compte tenu de notre attachement au physicalisme. Or, comme le dit Kim: "La sévérité d'un problème philosophique dépend de deux questions connexes: premièrement, à quel point sommes-nous attachés aux thèses qui donnent lieu au conflit présumé? Deuxièmement, à quel point est-il facile ou difficile de concilier de manière acceptable les thèses en litige?"[60] Pour le problème de la causalité mentale, les thèses en litige sont le dualisme des propriétés et le physicalisme. Et ce que fait l'argument de la survenance, c'est insister sur notre conviction en l'une des thèses, le physicalisme, pour nous amener à conclure à son incompatibilité avec l'autre, le dualisme des propriétés. À moins de découvrir une erreur dans le raisonnement de Kim, la seule alternative restante est de reconsidérer notre attachement aux différentes thèses qui entrent en conflit.

Une fois écartées les approches se situant à l'intérieur du physicalisme non-réductif, seules quatre possibilités subsistent. En premier lieu, il est possible d'adopter le dualisme des substances, à la condition toutefois d'abandonner le physicalisme, c'est-à-dire la thèse de la fermeture causale du monde physique et l'idée que le mental survient sur le physique. Peu de philosophes seront prêts à franchir ce pas. En effet, le physicalisme est une thèse de premier plan dans notre ontologie et le dualisme des substances n'a jamais pu régler le problème de la causalité mentale. Les trois autres approches, elles, conservent le physicalisme. En second lieu, nous trouvons l'épiphénoménisme, qui a le mérite de conserver non seulement le physicalisme, mais aussi le dualisme des propriétés. Le prix à payer, cependant, est l'idée de la causalité mentale elle-même: l'épiphénoménisme n'accorde aucun pouvoir causal aux propriétés mentales. Restent donc les approches qui rejettent le dualisme des propriétés. À ce titre se trouve ainsi la troisième approche, l'éliminativisme, qui nie purement et simplement

---

[60] *"The seriousness of a philosophical problem therefore depends on two related questions: First, how deep is our attachment to the assumptions and commitments that give rise to the apparent conflict? Second, how easy or difficult is it to bring the conflicting assumptions into an acceptable reconciliation?"* Kim (1998), p. 30.

l'idée du mental et favorise le recours à des explications strictement matérialistes pour rendre compte de nos actions. Entre l'épiphénoménisme et l'éliminativisme, le réductionnisme fait presque figure de compromis: le mental n'est pas éliminé comme dans le cas de l'éliminativisme et le mental n'est pas non plus dénué de pouvoirs causaux. C'est sans doute un peu ainsi que le voit Kim, d'ailleurs. Mais il ne faut pas s'y laisser prendre: le réductionnisme (peu importe la variante) s'oppose au dualisme des propriétés. Par conséquent, le 'mental' dont on parle est totalement physique. Les concepts de la psychologie (tant scientifique que 'populaire') ne continuent d'exister qu'en tant qu'ils réfèrent à des propriétés physiques, puisque le mental se réduit au physique. Parler de causalité mentale dans ces conditions relève donc quelque peu de l'exagération: s'il y a bien causalité, elle n'a de mental que le nom. La préservation de la théorie du mental que nous offre le réductionnisme ne constitue donc qu'un avantage mineur en regard des pertes qu'il nous force à accepter. Sans compter que nous ne disposons toujours pas de modèle réductif réellement satisfaisant: le modèle de la réduction fonctionnelle de Kim ne réussit pas à rendre compte des qualia. De plus, le pouvoir explicatif que Kim accorde à l'identification n'est pas vraiment supérieur, tout compte fait, à celui que peut avoir la survenance.

Reste bien sûr l'émergentisme, porteur de grandes promesses aussi bien que de grands périls. Il est difficile d'évaluer la valeur de l'émergence à ce stade. Cette relation pourrait éventuellement règler tous nos problèmes, mais la version traditionnelle de l'émergentisme a toujours eu un côté un peu magique. L'émergence apparaît souvent comme une réponse facile pour expliquer les phénomènes que l'on ne comprend pas. Tous les exemples d'émergence que nous ont offert les premiers partisans de l'émergence au début du XXe siècle se sont révélés n'être que des cas de propriétés résultantes. On ne peut bien sûr pas en inférer que les cas qu'aujourd'hui nous croyons être émergents n'en sont pas, mais cela nous force certainement à être prudent. Tant que des modèles plus sérieux d'émergence, comme celui de Humpreys, n'auront pas fait leur preuve, mieux vaut sans doute ne pas considérer l'émergence comme une approche offrant une véritable solution.

Nous voici donc face à un dilemme: si nous acceptons l'existence de propriétés mentales irréductibles, les qualia, alors tous nos efforts pour réduire les propriétés mentales intentionnelles sont vains. Car tant qu'à renoncer à rendre compte d'une partie de nos états mentaux, autant renoncer à tous les expliquer et ainsi préserver l'idée de la causalité mentale qui nous est si chère. Ou alors nous pouvons rejeter l'existence des qualia, mais dans ce cas nous devons sacrifier l'idée de la causalité mentale et la moitié de nos états mentaux pour finir avec une métaphysique de l'esprit cohérente certes, mais qui parle à peine du mental, la moitié étant composée d'éliminativisme et l'autre d'un réductionnisme insatisfaisant.

Ce n'est pas que je préconise une autre alternative. L'argument de la survenance de Kim me semble solide et le dualisme des substances n'est pas une solution à nos problèmes. Quant à l'épiphénoménisme, je suis d'accord avec Kim qu'il ne fait aucun sens de parler de l'existence de quelque chose n'ayant aucune influence sur rien, et qui est pour cette raison indétectable non seulement en pratique, mais aussi en principe. Le débat actuel sur le statut du mental serait alors dans la même catégorie que le débat sur l'éther, cette substance qui était supposée combler le vide de l'espace, qui a animé longtemps les physiciens. Ne reste donc que l'éliminativisme et le réductionnisme, et bien que la marge soit mince entre les deux, le réductionnisme reste préférable à l'autre. Je crois d'ailleurs que la science, en particulier les neurosciences, suivront cette tendance. Mais il est clair pour moi que toute réduction se fait au prix de l'oblitération de certaines questions philosophiques. Faire une identification, c'est rendre illégitime la question 'pourquoi' plutôt que d'y répondre.

Références bibliographiques

Block, N.,

1980, "What is Functionalism?" Readings in Philosophy of Psychology, 1, N. Block, dir., Cambridge: Harvard University Press, 171-184.

Block, N., Flanagan, O., Güzeldere, G., dirs.,

1997, The Nature of Consciousness: Philosophical Debates, Cambridge: MIT Press.

Churchland, P. M.,

1981, "Eliminative Materialism and the Propositional Attitudes", Journal of Philosophy, 78, 67-90.

1985, "Reduction, Qualia and the Direct Introspection of Brain States", Journal of Philosophy, 82, 8-28.

David, M.,

1997, "Kim's Functionalism", Philosophical Perspectives, 11, 133-148.

Davidson, D.,

1970, "Mental Events", repris dans Davidson (1980), Essays on Actions and Events, Oxford: Clarendon Press.

1993, "Thinking Causes", dans J. Heil et A. Mele, dirs. (1993) 3-17.

Dennett, D. C.,

1991, Consciousness Explained, Boston: Little Brown.

Dreske, F.,

1991, "Dreske's Replies" dans Dreske and His Critics, B. McLaughlin, dirs., Oxford: Blackwell, 180-221.

Feigl, H.,

1958, The "Mental" and the "Physical", Minneapolis: University of Minnesota Press.

Guttenplan, S., dir.,

1994, A Companion of the Philosophy of Mind, Oxford: Blackwell.

Hele, J. et Mele, A., dirs.,

1993, Mental Causation, Oxford: Clarendon.

Horgan, T.,

1997, "Kim on Mental Causation and Causal Exclusion", Philosophical Perspectives, 11, 165-184.

Humphreys, P.,

1996, "Aspects of Emergence", Philosophical Topics, vol. 24, no. 1, 53-70.

1997a, "Emergence, Not Supervenience", Philosophy of Science, 64, S337-S345.

1997b, "How Properties Emerge", Philosophy of Science, 64, 1-17.

Jackson, F.,

1982, "Epiphenomenal Qualia", American Philosophical Quarterly, 127-136.

1986, "What Mary Didn't Know" Journal of Philosophy, 291-295.

Kim, J.,

1982, "Psychophysical Supervenience", Philosophical Studies, 41, 51-70, aussi dans Kim (1993), 175-193.

1984, "Concepts of Supervenience", Philosophy and Phenomenological Research, 45, 153-176, aussi dans Kim (1993), 53-78.

1987, "'Strong' and 'Global' Supervenience Revisited", Philosophy and Phenomenological Research, 48, 315-326, aussi dans Kim (1993), 79-91.

1989a, "Mechanism, Purpose, and Explanatory Exclusion", Philosophical Perspectives, 3, 77-108, aussi dans Kim (1993), 237-264.

1989b, "The Myth of Nonreductive Materialism", Proceedings and Addresses of the American Philosophical Association, 63, 31-47, aussi dans Kim (1993), 265-284.

1990, "Supervenience as a Philosophical Concept", Metaphilosophy, 21, 1-27, aussi dans Kim (1993), 131-160.

1992, "Multiple Realization and the Metaphysics of Reduction", Philosophy and Phenomenological Research, 52, 1-26, aussi dans Kim (1993), 309-335.

1993a, "The Nonreductivist's Troubles with Mental Causation", Mental Causation, eds. Heil and Mele, Oxford: Oxford University Press, 189-210, aussi dans Kim (1993), 336-357.

1993b, "Postcripts on Mental Causation" dans Kim (1993), 358-367.

1993c, Supervenience and Mind: Selected Philosophical Essays, Cambridge University Press.

1996, Philosophy of Mind, Boulder: Westview Press.

1998, Mind in a Physical World, Cambridge: MIT Press.

----, "Making Sense of Emergence", à paraître dans Philosophical Studies, 1-33.

Levine, J.,

1983, "Materialism and Qualia: The Explanatory Gap", Pacific Philosophical Quarterly, 64, 354-361.

Lewis, D.

1983, "Postscript to 'Mad Pain and Martian Pain'" dans Philosophical Papers, 1, Oxford: Oxford University Press.

Macdonald, C. et Macdonald, G., dirs.,

1995, Philosophy of Psychology: Debates on Psychological Explanation, Cambridge: Blackwell.

McGinn, C.,

1989, "Can We Solve the Mind-Body Problem?", Mind, 98, 391, 349-366, aussi dans Block et *al.* (1997).

Marras, A.,

1998, "Kim's Principle of Explanatory Exclusion", Australasian Journal of Philosophy, 76, 3, 439-451.

Nagel, E.,

1961, The Structure of Science, New York: Harcourt, Brace.

Nagel, T.,

1974, "What Is It Like to Be a Bat?", Philosophical Review, 83, 4, 435-450, aussi dans Block et *al.* (1997).

Nemirow, L.,

1980, "Review of T. Nagel *Mortal Questions*", Philosophical Review, 475-476.

Searle, J. R.,

1992, The Rediscovery of the Mind, Cambridge: MIT Press.

Smart, J. J. C.,

1959, "Sensation and Brain Processes". Philosophical Review, 68, 141-156.

Van Gulick, R.,

1993, "Who's in Charge Here? And Who's Doing All the Work?" dans J. Heil et A. Mele dirs. (1993), p. 233-256.

1997, "Understanding the Phenomenal Mind: Are We All Just Armadillos?", dans Block et *al.* (1997). Part I: 559-566 et Part II: 435-442.