

Le fonctionnalisme non réductionniste et la réalisabilité multiple

L'irréductibilité mise à l'épreuve

FRÉDÉRIC DUBOIS, *Université Laval*

RÉSUMÉ: L'argument de la réalisabilité multiple est une pierre angulaire de bien des théories physicalistes non réductionnistes, plus particulièrement celles du fonctionnalisme d'Hilary Putnam et de Jerry Fodor. Si l'argument de la réalisabilité multiple prétend pouvoir bloquer la réduction des états mentaux à de simples états physiques, nous pouvons alors aussi espérer que l'argument peut résister aux défis de la causalité mentale lancés par Jaegwon Kim, à savoir les problèmes de l'épiphénoménisme, de l'exclusion causale et de la surdétermination. Notre thèse vise à montrer les difficultés qu'éprouvent la réalisabilité multiple et le fonctionnalisme à pouvoir, au final, éviter les considérations kimiennes sur la causalité mentale. Nous montrerons que la réalisabilité multiple peut mieux être comprise comme un argument épistémologique, mais dès lors avec de lourdes conséquences ontologiques.

Introduction

Les recherches contemporaines menées par les neurosciences cognitives laissent fortement suggérer une corrélation directe entre les phénomènes physiques cérébraux et les processus mentaux. En effet, le perfectionnement des outils d'imagerie médicale nous permet aujourd'hui d'observer l'impact des lésions cérébrales sur différentes aptitudes mentales, de la mémoire jusqu'aux traits psychologiques d'un individu. Nous savons également que les psychotropes, l'alcool, les médicaments et autres substances entraînent des transformations chimiques dans le cerveau, ayant à leur tour des impacts significatifs sur les humeurs et les émotions. Il y aurait donc des liens causaux serrés du physique vers le mental¹. En fait, les théoriciens physicalistes

en général, mais particulièrement les réductionnistes, avancent même que nous expérimentons des états mentaux seulement parce que notre cerveau se trouve dans un certain état chimico-physique, faisant ainsi de la vie mentale un ensemble d'états physiques dépendant du cerveau ou du système nerveux central². La *thèse de la corrélation entre le mental et le physique* (CMP)³ peut dans cette lignée être formalisée de la sorte :

Pour chaque événement mental de type M qui a lieu dans un organisme *o*, il existe un état physico-cérébral de type P tel que l'événement de type M a lieu dans *o* à un temps *t* si et seulement si (*ssi*) un événement P a lieu dans *o* à *t*.

Suivant la CMP, nous pourrions dire qu'un événement mental doit avoir lieu dans un organisme (ou structure-type) et posséder un corrélat neural à la fois suffisant et nécessaire à son occurrence. Il s'ensuit également qu'un état mental M doit nomologiquement être en corrélation avec un état physique P. Par exemple, si l'on prend un état mental tel que la douleur et son corrélat physique l'excitation des fibres C (efC), la CMP affirmera que lorsqu'une douleur survient, un certain état physique, soit efC, survient aussi. L'expérience de la douleur et le phénomène physique efC sont ainsi mis en corrélation de manière nomique puisque le fait que nous éprouvons la douleur *ssi* efC démontre une régularité, et non une simple cooccurrence accidentelle. De plus, si pour chaque fois que M se produit P doit également se produire, il s'ensuit logiquement que la moindre modification au sein de la vie mentale doit être accompagnée d'une modification au sein des propriétés physiques.

Dès lors, nous pouvons nous demander : quel rôle causal reste-t-il aux états mentaux ? En effet, la CMP, que nous associons ici à une théorie de l'identité (bien que nous puissions aussi l'associer de manière plus générale à d'autres théories physicalistes), semble ne faire des états mentaux, telle la douleur, qu'une simple reformulation en termes psychologiques d'un phénomène physique efC déjà suffisant en lui-même pour expliquer le phénomène à l'œuvre dans le cerveau. En bref, la théorie de l'identité soutient que « P = M » (d'autres théories physicalistes adopteraient plutôt la posture selon laquelle M survient sur P), ce qui, au final, nous laisse perplexes

sur la pertinence causale de M. Aurions-nous même besoin de deux expressions différentes pour décrire un même phénomène dont l'explication physique suffit à rendre compte? Certes, l'usage de deux expressions distinctes peut demeurer pertinent du point de vue pragmatique. En effet, de la même manière que le terme «eau» et son équivalent chimique « H_2O » réfèrent à la même substance, donc à une seule et même chose, le langage commun tend à les dissocier, faisant de l'un un terme générique et de l'autre une expression scientifique. Ils n'ont au final qu'une fonction pragmatique dans la langue se résumant à une différence au niveau du discours. De la même manière, si l'on accepte la thèse réductionniste « $P = M$ », il ne resterait plus qu'à dire que la douleur relève de la psychologie populaire, ce qu'un physicaliste réductionniste dira n'être en fait qu'une efC^4 .

Malgré cette intuition d'identité forte suggérée par les neurosciences, le postulat « $P = M$ », ou encore (bien que non équivalent) que M survient sur P, nous conduit pourtant à une série de problèmes d'ordre logique et philosophique: (a) soit les phénomènes mentaux ne seraient alors que des épiphénomènes, en bref inutiles (problème de l'épiphénoménisme); (b) soit, en vertu du principe de la fermeture causale du monde physique (ou principe de complétude⁵) voulant que tous phénomènes physiques doivent nécessairement avoir une cause physique suffisante pour expliquer ses effets, les phénomènes mentaux, s'ils sont autres que physiques, semblent alors être exclus de la causalité (problème de l'exclusion causale⁶), et si les phénomènes mentaux sont en fait physiques, ils sont donc réductibles et sûrement, à nouveau, épiphénoménaux; (c) soit M joue bel et bien un rôle causal, mais dès lors nous assisterions à un cas de surdétermination puisque P, déjà suffisant pour causer le phénomène donné, se voit ajouter une cause mentale supplémentaire (problème de la surdétermination). Cela nous reconduirait une fois de plus à remettre en doute la pertinence causale de M.

Pourtant, sauf si l'on défend une posture épiphénoménaliste, il semble y avoir un *desideratum* que les propriétés mentales conservent certains pouvoirs causaux, c'est-à-dire qu'elles soient aptes à causer d'autres événements. En faire autrement reviendrait à remettre en

doute nos intuitions sur la manière dont fonctionne la causalité mentale dans la vie de tous les jours. Il semble en effet demeurer un *quelque chose de plus* qu'une série de processus neurochimiques dans l'exécution de nos actions quotidiennes. Nous voudrions alors que ce qui cause notre action, par exemple aller à la cuisine se chercher une tasse de thé, soit plus qu'une cause déterminée par nos neurones. Nous voudrions ainsi que l'une des causes tout aussi pertinentes puisse être le désir du thé, et la croyance qu'en mettant notre corps en mouvement vers la cuisine ce désir puisse se réaliser. En résumé, concevoir que notre monde peut tout aussi bien être sans nos états mentaux (ou qu'ils sont tout simplement inutiles) revient, selon l'expression de Jerry Fodor, à considérer la « fin du monde »⁷. En effet, la victoire des théories de l'identité marquerait la défaite de la psychologie et rendrait caduque la conception que nous avons de nous-mêmes en tant qu'agents libres aptes, par nos désirs et nos croyances, à changer le monde. Il ne resterait plus qu'à faire de la place pour les neurosciences qui détermineraient à elles seules la localisation de nos phénomènes cérébraux déterminés physiquement et jadis fautivement interprétés comme étant « mentaux »⁸.

Bon nombre de philosophes se sont donc penchés sur la question et ont essayé de réconcilier le physicalisme avec une approche qui permettrait à la fois de conserver nos intuitions sur la causalité mentale, puis de sauvegarder le rôle des sciences particulières comme la psychologie. L'approche du physicalisme non réductionniste par le fonctionnalisme d'Hilary Putnam s'est avérée être l'une des plus fructueuses à cet égard⁹. Au cœur de la position de Putnam se trouve l'argument de la réalisabilité multiple (RM) voulant qu'une propriété mentale puisse être réalisée, instanciée ou implémentée par une multiplicité de propriétés physiques. Tout effort de postuler une théorie de l'identité stricte serait alors mis à mal du simple fait que la réduction « $P = M$ » ne peut, dans tous les mondes possibles, être vraie. Le fait qu'il soit concevable et possible que M soit réalisé parfois par P_1 , parfois par P_2 (et possiblement par un plus grand nombre de propriétés P_n), mais que P_1 et P_2 ne partagent pas les mêmes propriétés, est suffisant pour montrer que « $P = M$ » est contingent et non pas nécessaire.

Bien que nous concédions qu'il existe une pluralité d'approches fonctionnalistes en philosophie de l'esprit, qu'elles soient parfois non réductionnistes (par exemple le fonctionnalisme du rôle) et d'autres fois réductionnistes (par exemple le fonctionnalisme de l'occupant du rôle)¹⁰, notre présente analyse se réduira au fonctionnalisme causal non réductionniste, également dit computationnaliste, tel que développé à partir des années 60 par Putnam et Fodor. Ce fonctionnalisme non réductionniste se veut néanmoins physicaliste et cherche à répondre aux difficultés posées par la causalité mentale en faisant appel à l'argument de la RM, longtemps accepté et assez peu critiqué avant les travaux de Jaegwon Kim dans les années 90. Après avoir mieux exposé en détail les rouages de la RM (section 1), nous montrerons comment le fonctionnalisme peut chercher à réconcilier la causalité mentale avec l'intuition que nos états mentaux ont tous pour base des états physiques. Nous revisiterons par la suite (section 2) les arguments sur lesquels repose la RM (principalement l'argument de la concevabilité et l'argument empirique) afin de voir s'ils peuvent résister aux contre-arguments apportés par le physicalisme réductionniste. Ensuite, nous reconduirons (section 3) notre réflexion sur la causalité fonctionnelle à partir des deux critères nécessaires à la RM, soit la différenciation des propriétés réalisatrices et l'identité des propriétés réalisées. Un dernier examen (section 4) du fonctionnalisme non réductionniste et de la RM posera finalement la question des pouvoirs causaux à la lumière des critères de différenciation et d'identité préalablement postulés. Nous ne chercherons pas dans ce travail à prouver hors de tout doute que les propriétés mentales sont véridiquement réalisées de manière multiple ou non, mais plutôt à voir les conséquences de telles réalisations sur l'interaction causale du mental et du physique. Nos conclusions suivront celles de Kim¹¹, c'est-à-dire que nous en viendrons à nous demander si non seulement le fonctionnalisme et la RM peuvent en effet éviter le réductionnisme, mais aussi si tous deux ne nous conduiraient pas plutôt vers une réduction multiple. Si cette conclusion s'avère juste, les conséquences seraient lourdes pour la RM au niveau ontologique dans la mesure où nous n'assisterions

plus à une simple corrélation stricte entre P et M¹², mais bien à une multitude de réductions possibles entre P₁ et M₁, P₂ et M₂... P_n et M_n.

1. La thèse de la réalisabilité multiple et le fonctionnalisme causal

La visée première de la RM telle que nous la retrouvons chez Putnam¹³ est celle de réfuter la théorie de l'identité des types¹⁴. Putnam voulait ainsi montrer avec la RM que, contrairement à ce qu'affirment les théoriciens de l'identité des types, les états mentaux (par exemple la douleur) ne sont pas identiques, et donc ne peuvent pas être réduits à des états physiques (par exemple l'excitation des fibres C). L'argument avancé par Putnam est qu'il est possible, donc concevable, qu'un état mental tel que la douleur se manifeste dans différents organismes bien que les structures physico-chimiques de ces mêmes organismes soient différentes de ce que l'on retrouve chez les humains et les mammifères en général. Ces organismes feraient tout de même l'expérience de la douleur au même titre qu'un organisme ayant des fibres C, bien que le déclenchement de l'état mental se ferait par l'entremise, disons, des fibres Q. Donc, la RM conçoit la possibilité que M soit en fait instancié dans une structure neurobiologique différente, non plus par P, mais peut-être par P₁, P₂... P_n. Les propriétés neurophysiologiques réaliseraient alors les états mentaux, sans pour autant s'y réduire dès lors que l'on accepte que les états psychologiques individuels puissent être réalisés par une multiplicité de propriétés physiques différentes. Nous pourrions aussi conclure que ni les fibres C, ni les fibres Q chez une créature peut-être extraterrestre ou encore un robot, ne sont nécessaires en elles-mêmes pour que cette créature ressente de la douleur, puisque la douleur pourrait être observée en l'absence de l'une ou de l'autre de ces fibres. Du même coup, cela démontre que nous ne pouvons pas formuler une relation d'identité stricte entre l'excitation des fibres C (état physique P) et la douleur (état mental M) tel que précédemment conclu à partir de la CMP. En effet, si «P = M», mais que M peut également être réalisé tant par P₁ que par P₂ (donc que «P₁ = M» et «P₂ = M»), cela voudrait alors dire que «P₁ = P₂». En vertu de la loi des indiscernables de Leibniz, cela reviendrait à dire que P₁ et P₂ devraient partager les mêmes propriétés. Or, cela n'étant pas le cas,

on peut avancer que P_1 et P_2 ne sont pas identiques, bloquant ainsi la réduction de M à une propriété physique de base individuelle. Au final, la RM avance que les propriétés mentales n'ont pas de propriétés physiques nomologiquement coextensives du moment qu'un type P s'avère réalisable de manière multiple ($P_1, P_2 \dots P_n$). L'argument se résume alors ainsi :

- (1) Tous types mentaux sont réalisables de manière multiple par différents types physiques.
- (2) Si un type mental donné est réalisable de manière multiple par différents types physiques, alors il ne peut être identique à aucun type physique spécifique. Donc,
- (3) Aucun type mental n'est identique à un type physique spécifique¹⁵.

Si l'on adopte la thèse de la RM, il s'ensuit que les événements mentaux semblent moins contraints par les structures neurales (et neuronales) qui les réalisent. C'est ainsi dire qu'ils fonctionnent sensiblement de la même manière que des concepts d'artefacts. Par exemple, le concept de moteur, en soi, ne nous dit rien sur le mécanisme physique concret qui le réalise. En effet, un moteur peut fonctionner à l'essence, à l'électricité, à la vapeur ou encore même avec des composantes hybrides. Le concept de l'artefact moteur ne nous renseigne pas plus sur le nombre de pistons que possède éventuellement un moteur, que sur le fait qu'il est à combustion interne, à explosion ou bien à allumage commandé. Nous remarquons donc que certaines propriétés (couleur, composantes, etc.) n'ont pas d'importance pour la description du concept d'un artefact. Ce qui définit en fait un objet tel qu'un moteur, c'est la description du travail qu'il accomplit, donc son rôle causal, à savoir permettre d'effectuer un mouvement en transformant une énergie mécanique en une énergie de nature différente. Nous arrivons avec ces considérations à une conception fonctionnaliste des interactions du mental et du physique. L'approche fonctionnaliste cherche alors à définir les états mentaux de la même manière que les concepts d'artefacts, c'est-à-dire non pas par la description des mécanismes qui les rendent possibles, mais bien par la fonction qu'ils remplissent. Une définition fonctionnaliste de la douleur serait alors ainsi retranscrite :

De la même manière que x est un moteur =_{def} x a une certaine propriété P telle que P permet à x de transformer une énergie en mouvement ; x est une douleur =_{def} x a une certaine propriété P telle que P permet à x de détecter des lésions tissulaires¹⁶.

Le fonctionnalisme fait alors appel aux propriétés fonctionnelles afin d'expliquer les phénomènes mentaux. F est une propriété fonctionnelle dans le cas où F peut être caractérisée par une définition de la forme suivante :

Pour une chose x , avoir F (ou être un F) =_{def} pour x , avoir une propriété P telle que $C(P)$, où $C(P)$ est une spécification de la tâche causale que P est censée remplir dans x .

À partir d'une définition fonctionnelle, couplée aux propriétés fonctionnelles causales, le fonctionnalisme conçoit la RM de manière à ce que

Une propriété P réalise F (c'est-à-dire que P est un réalisateur de F) dans un système S *ssi* $C(P)$, à savoir si P correspond à la spécification causale C dans S ¹⁷.

Cela revient plus simplement à dire que P effectue la tâche causale spécifiée dans le système S . Dans ce cas, on définit toujours l'état mental par la fonction effective réalisée par P et non pas par la structure qui le compose. Pour reprendre l'exemple récurrent de la douleur, sa fonction n'est évidemment pas efC , mais plutôt d'agir comme un détecteur de lésions tissulaires. L'occurrence d'une lésion sur le corps devrait être suffisante pour activer le mécanisme acheminant la douleur, ce qui entraînera un certain comportement de la part du sujet faisant l'expérience de la douleur. La fonction de la douleur sert ainsi d'intermédiaire causal entre une « entrée » (lésion tissulaire) et une « sortie » (grimace, gémissement, etc.). Lorsqu'on joint cette conception des propriétés mentales avec la RM, nous arrivons également à la conclusion que pour toute propriété M , il y a une variété de propriétés physiques P (P_1, P_2, \dots, P_n), chacune répondant aux spécifications C du système S et qui instancient la propriété M .

De même, nous pouvons dire qu'une propriété fonctionnelle F est identique à, disons, G , *ssi* le rôle causal attribué à F est le même que celui attribué à G . F et G sont alors, selon la terminologie de Putnam, isomorphes fonctionnellement. Ainsi que Putnam le

définit, «deux systèmes sont fonctionnellement isomorphes s'il y a une correspondance entre les états de l'un et les états de l'autre qui préserve les relations fonctionnelles»¹⁸. De cette manière, on pourrait concevoir qu'un humain et un extraterrestre ne partageant pas les mêmes structures neurophysiologiques seraient néanmoins isomorphes fonctionnellement au niveau de la douleur *ssi* les fonctions causales instanciées par efC chez l'humain sont exactement le reflet des fonctions causales implémentées par efQ chez l'extraterrestre. Nous reviendrons aux sections 3 et 4 sur les conséquences d'une telle conception des pouvoirs causaux.

2. La réalisabilité multiple mise à l'épreuve

Bien que nous ne l'ayons pas encore observé scientifiquement, l'idée que $P_1, P_2 \dots P_n$ existent dans l'univers ou dans un autre monde possible à l'instar de P se fonde sur l'argument de la concevabilité. Cette stratégie était déjà utilisée par Descartes dans sa réflexion sur la corrélation (ou plutôt la dichotomisation) du corps et de l'esprit. Descartes avançait que si l'on pouvait clairement imaginer une chose x pouvant exister séparément de y , clairement x ne peut pas être y . Au sein d'une théorie de la RM, l'argument se retranscrit comme suit :

- (1) Il est clairement et distinctement concevable (pour reprendre la formule de Descartes) que différentes propriétés physiques puissent instancier les mêmes états mentaux ;
- (2) Ce qui est clairement et distinctement concevable est possible. D'où,
- (3) Il est possible que différentes propriétés physiques instancient les mêmes états mentaux ;
- (4) S'il est possible que différentes propriétés physiques instancient les mêmes états mentaux, les états mentaux doivent être réalisables de manière multiple. Donc,
- (5) Les états mentaux sont réalisables de manière multiple.

Montrons tout d'abord comment l'argument pourrait être valide (ou non) à partir de la logique modale. S'il est concevable que les propriétés P_1 et P_2 puissent instancier une même propriété M , il doit exister un monde possible où P_1 réalise M (mais non P_2), un autre monde possible où P_2 réalise M (mais non P_1), et possiblement un

monde où sont présents à la fois P_1 et P_2 . Pour que nous puissions postuler une relation d'identité stricte à partir de la CMP, il faudrait plutôt qu'un même P soit le seul et unique corrélat de M dans tous les mondes possibles. L'équation « douleur = efC » devrait effectivement être, pour une théorie de l'identité, une vérité nécessaire, valide dans tous les mondes possibles. Cependant, si l'on en croit la RM, il semble tout à fait concevable, à partir de la prémisse (1), que différentes propriétés physiques instancient un même état mental. Cela revient à dire que la douleur pourrait possiblement exister sans la présence d'efC (grâce à efQ chez l'extraterrestre, par exemple). Si une telle possibilité existe, l'équation « douleur = efC » ne peut alors plus être une vérité nécessaire. En effet, dans la mesure où il est concevable que P_2 réalise M sans la nécessité de l'occurrence de P_1 , il s'ensuit que chaque type de P doit être contingent à la réalisation de M . Pour le répéter, la RM attaque donc la nécessité d'une identité stricte en montrant la contingence de P . En logique modale, une identité ne peut être contingente que si l'une des deux expressions « X » ou « Y » est ce que Saul A. Kripke nomme un « désignateur non rigide »¹⁹. Un désignateur rigide réfère à un objet qui dans tous les mondes possibles est nécessairement le même. Prenons l'exemple « Justin Trudeau = Le 23^e Premier Ministre du Canada ». « Justin Trudeau » ne peut être qu'un désignateur rigide. Certes, Justin pourrait dans un monde possible se nommer différemment. Le Justin du monde possible α pourrait aussi avoir les cheveux roux. Un autre Justin du monde possible β pourrait avoir les cheveux blonds. Ces différences de propriétés physiques n'empêchent pas qu'en toute tautologie « Justin = Justin » dans tous ces mondes possibles. Cela peut-il également être le cas pour l'expression « Le 23^e Premier Ministre du Canada » ? Nous pouvons facilement concevoir un monde où Thomas Mulcair aurait remporté les élections canadiennes et serait ainsi devenu le 23^e Premier Ministre du Canada. Cette dernière expression n'est ainsi pas vraie de Justin Trudeau dans tous les mondes possibles, et nous devons alors conclure que « Justin Trudeau = Le 23^e Premier Ministre du Canada » est une vérité contingente.

En recourant à la logique modale, il suffirait alors de montrer que soit « P » (par exemple « efC ») soit « M » (par exemple « douleur »)

sont des désignateurs non rigides afin de réfuter la nécessité de « $P = M$ ». En effet, si deux termes « X » et « Y » sont tous deux des désignateurs rigides, il s'ensuivrait que l'identité « $X = Y$ » serait nécessairement vraie, de la même manière que «eau = H_2O » est *nécessairement* vraie. Or, si soit «douleur» ou soit «efC» sont des désignateurs non rigides, «douleur = efC» ne sera pas une vérité nécessaire, mais au mieux une vérité contingente, valide dans certains mondes possibles ou pour certains organismes, tout comme de manière plus générale toutes les identités psychophysiques. Est-ce que «efC» ou «douleur» peuvent être considérés comme des désignateurs non rigides? Il ne semble pas évident d'affirmer que «efC» puisse être un désignateur non rigide. L'excitation des fibres C ne peut être un événement autre (par exemple, l'éruption du Vésuve ou la collision entre deux étoiles, etc.) que l'excitation des fibres C. Comme l'écrit Kim, «un monde dans lequel l'excitation des fibres C n'a jamais lieu est un monde où cet événement, l'excitation des fibres C, n'a pas lieu»²⁰. Certes, les fibres C telles que nous les connaissons pourraient être, structurellement, totalement différentes, à l'instar de Justin Trudeau- α aux cheveux roux. Néanmoins, que les propriétés physiques des fibres C varient d'un monde à l'autre n'enlève pas au fait qu'il est nécessaire qu'une efC soit une efC et rien d'autre. Parce que la RM stipule qu'autre chose qu'une efC peut nécessiter la douleur, elle peut imaginer, par exemple, des excitations de fibres Q coexistantes avec des excitations de fibres C dans un monde possible. Que cela soit possible ne remet toutefois pas en cause que «fibres C = fibres C- α ». Il demeurera plus tard à montrer ce que l'excitation des fibres C et l'excitation des fibres Q (efQ) ont en commun pour que nous puissions les associer toutes les deux à la douleur. Pour ce qui nous intéresse présentement, retenons la rigidité de l'expression «efC».

Maintenant que nous avons établi que «efC» est un désignateur rigide, une corrélation contingente pourrait subsister si nous montrons que «douleur» est alors un désignateur non rigide. Comme le remarque Kim,

Si vous êtes enclin à considérer le caractère douloureux de la douleur comme sa propriété définitionnelle essentielle, vous

direz que «douleur» désigne de manière rigide un événement ou un état pourvu de cette qualité douloureuse, et que cette expression désigne un événement de cette sorte dans tous les mondes possibles. Un monde dans lequel rien ne fait jamais mal est un monde sans douleur²¹.

Autant «efC» que «douleur» semblent ainsi être des désignateurs rigides. Du même coup, la corrélation de nécessité avancée par la CMP semble ainsi pouvoir tenir. Tout de même, en dépit de la logique modale et pour l'exercice de pensée, accordons encore la véracité de la prémisse (1), c'est-à-dire qu'il est en effet concevable que différentes propriétés physiques instancient les mêmes états mentaux. Bien que «efC = douleur» soit nécessaire, il semble en effet concevable qu'un extraterrestre puisse ressentir de la douleur alors que l'on retrouve chez lui des fibres Q plutôt que des fibres C. D'ores et déjà, l'argument de la concevabilité ne nous dit pas si la prémisse est en effet *véritablement* possible, mais qu'elle est simplement cohérente et imaginable. En d'autres mots, bien que la prémisse puisse être épistémologiquement admissible, et donc qu'il soit possible d'imaginer qu'un extraterrestre puisse ressentir la douleur par l'excitation des fibres Q (efQ), cela ne nous assure pas pour autant que cette configuration neurale *est*, ontologiquement, réalisée. Ainsi, nous ne savons tout simplement pas si la prémisse est *réellement* correcte, mais simplement qu'il est cohérent et concevable logiquement qu'efQ soit concrétisé quelque part dans l'univers ou dans un monde possible. Il ne va pas de soi non plus en se basant sur la logique modale qu'efQ puisse coexister avec efC tout en réalisant le même phénomène de douleur. Le problème majeur de l'argument de la concevabilité se situe donc au niveau du fait que la prémisse (2) amalgame le «réel» (ce qui *est*, métaphysiquement) et la «possibilité» en tant que simple possibilité épistémique. C'est ainsi dire que (2) ne peut pas servir de base de vérité à (3) dès lors que la prémisse (2) fait glisser la question épistémologique à une réalité métaphysique concrète. Le mieux que l'on puisse établir de la prémisse (1) est qu'il est possible, dans un sens épistémologique de possibilité, d'avoir une propriété P_1 (efC) et une propriété P_2 (efQ) qui réalisent toutes deux la douleur dans deux structures S

différentes. Or, cette possibilité n'implique pas qu'il y ait dans l'univers des fibres Q. Imaginons, par exemple, un monde possible où l'eau ne soit pas de composition H_2O . Cela ferait en sorte qu'au niveau conceptuel et du point de vue logique, l'eau et l' H_2O peuvent être considérés comme indépendants. En considérant l'exercice de pensée de ce monde possible, l'équation «eau $\neq H_2O$ » devient tout aussi épistémologiquement concevable et cohérente. Néanmoins, il est nécessaire dans la réalité que «eau = H_2O » (et que les deux termes de l'équation sont tous deux des désignateurs rigides), hormis s'il s'agit d'une substance différente²². De la même manière, si efC et efQ coexistent dans un monde possible, il est tout aussi cohérent d'affirmer que nous serions en présence de deux types d'événements distincts réalisant deux états peut-être similaires, mais tout de même différents.

Pour que la thèse non réductionniste de la RM soit valide, il faut montrer en deux temps que (I) les réalisateurs sont distincts ($P_1, P_2 \dots P_n$), puis que (II) leurs propriétés fonctionnelles sont identiques. S'il demeure bien concevable que efQ existe, il ne semble pas pour autant aussi évident de dire que «efQ = douleur» au même titre que «efC = douleur». Pour les théoriciens de l'identité, des recherches plus poussées en neurosciences doivent être faites avant de pouvoir conclure, hors de tout doute, qu'efQ existe de la même manière qu'efC. Il faudrait alors démontrer empiriquement que la corrélation stricte entre M et P_1 est identique à celle entre M et P_2 , et non pas simplement la manifestation d'un phénomène semblable.

Les défenseurs de la RM ont cependant déjà certains outils théoriques leur permettant de valider leur argument empiriquement. La biologie évolutive, par exemple, nous donne déjà de bonnes raisons de croire en la possibilité de découvrir de la vie extraterrestre ayant évolué en fonction de conditions environnementales largement différentes de celles de la Terre. La rencontre d'une seule de ces espèces aurait de fortes chances d'appuyer la thèse de la RM. Or, il y a une forte probabilité que des créatures munies de structures neurobiologiques plus ou moins différentes présentent des états mentaux plus ou moins similaires (ou identiques) aux nôtres. Du point de vue de la biologie évolutive, la capacité à éprouver des états

psychologiques semble en effet être d'un certain avantage sélectif. On considère ainsi la capacité à ressentir la douleur comme un avantage au point de vue de la survie, puisque la douleur peut nous renseigner et prévenir contre d'éventuels dangers. La douleur peut donc aussi être conçue en fonction de son utilité à éloigner les menaces. En considérant l'avantage évolutif des états mentaux, et du fait que l'évolution biologique dépend grandement de l'environnement, on peut ainsi supposer qu'une vie extraterrestre aurait aussi développé, pour les mêmes avantages sélectifs, une vie mentale similaire (ou identique) à la nôtre. Puisque nous avons de bonnes raisons de croire que les environnements extraterrestres sont considérablement différents des conditions terrestres sur plusieurs variables, il en découle que les structures neurophysiologiques des extraterrestres devraient également présenter des différences considérables. Si tel est le cas, et que ces créatures extraterrestres peuvent aussi éprouver des états mentaux tels que la douleur, cela devrait être suffisant pour appuyer la thèse de la RM. Comme le souligne le paradoxe de Fermi²³, la réponse demeure cependant toujours indéterminée quant à savoir s'il est bel et bien possible de rencontrer ces extraterrestres et ainsi compléter nos recherches empiriques. L'argument de la biologie évolutive concernant la RM repose donc principalement sur l'induction, sans pour autant offrir à ce stade de quelconques preuves empiriques décisives²⁴.

Un deuxième argument empirique, tiré cette fois des neurosciences, peut toutefois être avancé afin d'appuyer la thèse de la RM. Les défenseurs de la RM peuvent faire appel à nos connaissances du phénomène de la plasticité cérébrale, déjà bien connu à l'époque où Putnam et Fodor développent leur argument de la RM. En résumé, les recherches sur le cerveau ont montré que les neurones ont la capacité, lorsque les structures neuronales sont endommagées, de prendre la relève sur les fonctions qui sont normalement remplies par les cellules avoisinantes. Les neurones qui sont réaffectés vers une autre tâche cessent donc d'être affectés à leur fonction initiale. Nous nous retrouvons alors devant une situation où à t_1 , P réalisait M_1 , puis à t_2 , P réalise M_2 . La réduction « $P = M$ » ne peut, encore

une fois, être considérée comme nécessaire du moment où ce même P pourrait réaliser $M_1, M_2 \dots M_n$.

Pourtant, en dépit de ce qu'avance la RM, les études scientifiques suggèrent plutôt que la plasticité neuronale est systématique, c'est-à-dire qu'elle suit une progression régulière (nomique) suite aux dommages des structures cérébrales. John Bickle²⁵ a d'ailleurs mis l'accent sur ce point, montrant que les neurosciences supposent *de facto* une continuité systématique sous-jacente aux mécanismes cérébraux. Certes, nos connaissances sur les mécanismes neuronaux sont encore aujourd'hui partielles. Les observations des neurosciences contemporaines suggèrent déjà néanmoins que la plasticité cérébrale n'arrive pas sans cause physique nomologiquement explicable. Un autre point problématique pour la RM est que l'on observe un affaiblissement des fonctions cérébrales suite au phénomène de plasticité cérébrale, alors que la qualité et la quantité des actions se voient grandement réduites chez les patients ayant subi des dommages au cortex cérébral, même avec le relais des neurones adjacents. Si la qualité et la quantité des fonctions semblent ainsi différer, pouvons-nous toujours postuler une RM? Les recherches menées par Jitendra Sharma, Alessandra Angelucci et Mriganka Sur²⁶ sur les furets peuvent nous donner un autre bon aperçu de la problématique sous-jacente. Ces recherches ont permis de trouver un moyen de rebrancher le cerveau de furets afin que l'information normalement opérée par le cortex visuel soit redirigée et opérée par le cortex auditif. En fonction des différences qui séparent les mécanismes du cortex visuel et du cortex auditif, cette expérience devrait nous convaincre qu'une propriété mentale peut en effet, comme l'avance la RM, être réalisée par différentes structures neurobiologiques. Lawrence A. Shapiro remarque qu'en effet le cortex auditif et le cortex visuel chez ces furets présentent des ressemblances quant à leur fonction respective, c'est-à-dire qu'ils reproduisent par exemple tous deux une carte spatiale en deux dimensions et que les cellules du cortex auditif des furets semblent, comme le laissait entendre le phénomène de la plasticité cérébrale, se transformer afin de reprendre la fonction normalement remplie par les neurones du cortex visuel²⁷. Cependant, les propriétés visuelles

des furets rebranchés ne semblent pas, comme c'était le cas avec la plasticité cérébrale, être le reflet exact de ceux des furets normaux quant à la qualité et la quantité des fonctions remplies. En effet, leur acuité semble nettement diminuée, ce qui entraîne une défaillance au niveau fonctionnel. Si leurs fonctions respectives diffèrent, que ce soit seulement de quelque peu, le critère d'identité (II) postulé par la RM ne semble pas être respecté. Cela pourrait aussi démontrer que même s'il est possible qu'une structure S puisse instancier une douleur M à partir d'efC, et que S_2 réalise également ce qui semble être une douleur à partir d'efQ, nous aurions tout autant de bonnes raisons d'induire que leurs fonctions respectives diffèrent qualitativement et quantitativement.

3. L'isomorphisme et la différenciation des propriétés

Selon Putnam, «si nous pouvons trouver un seul prédicat psychologique qui peut être clairement appliqué à la fois à un mammifère et à une pieuvre (par exemple “avoir faim”), mais dont le “corrélat” physico-chimique est différent dans les deux cas, la théorie de l'état cérébral s'effondre²⁸». Si l'on suit l'affirmation de Putnam selon laquelle les corrélats physico-chimiques des mammifères sont différents de ceux des pieuvres (ce que la science montre bel et bien), mais que l'on adopte en même temps la RM, il s'ensuit que ces deux différentes structures types devraient réaliser les *mêmes* états mentaux. Nous avons déjà démontré certaines difficultés que rencontre ce postulat de différenciation des propriétés (I) instanciant de nouvelles propriétés identiques (II). Nous chercherons ici à approfondir les difficultés déjà énoncées afin de rendre compte de l'étendue des conséquences qu'elles représentent pour la causalité mentale ainsi que l'ontologie qui en découlerait. Prenons pour cette section « N_1 » pour représenter la structure neuronale des mammifères et « N_2 » pour celle des pieuvres. Afin de démontrer la RM, il faut donc répondre aux critères de différenciation (DIF) des propriétés physiques de N_1 et N_2 tout en montrant que le M qu'ils réalisent est identique (IDE).

DIF. Les structures neuronales des mammifères (N_1) et des pieuvres (N_2) sont différentes.

IDE. L'état mental M réalisé par N_1 et N_2 est exactement identique.

Accordons-nous que postuler une identité des états mentaux est une thèse plus forte que de postuler une similarité. Or, la RM prend comme prémisses que N_1 et N_2 peuvent réaliser M , et non pas M_1 et M_2 , deux états mentaux peut-être similaires, mais néanmoins différents. Sur ce point, il faut cependant s'assurer d'être bien clair sur ce qu'entend de manière générale la RM par « identité ». Dire que les angles d'un triangle équilatéral sont identiques ne signifie pas que ces trois angles ne font qu'un, mais bien qu'ils partagent tous les trois des propriétés identiques (soit, ici, la mesure de l'angle de 60°). Dans la même lignée, Kate et Delphine pourraient toutes deux posséder un exemplaire d'*Être et temps* de Martin Heidegger. Or, cela implique qu'elles ont deux exemplaires d'un même livre, qui n'ont pas plus à être identiques au point de vue de toutes leurs propriétés physiques. Peut-être que Kate n'est finalement pas friande de la philosophie heideggérienne, et ainsi son livre, qu'elle n'a jamais ouvert, demeure en parfaite condition, quoique poussiéreux, dans sa bibliothèque. Delphine, au contraire, pourrait s'être trouvée une passion inexorable pour la phénoménologie herméneutique. Son exemplaire serait donc parsemé de notes et les pages pourraient être usées de longues heures de consultation. L'identité des angles du triangle équilatéral et celle des exemplaires de l'ouvrage de Heidegger sont ici comprises dans son sens *large*.

Lorsqu'une chose X est identique à une chose Y au sens strict, nous devons être en présence d'une seule et même chose (par exemple, l'eau est identique à sa composition chimique H_2O). Certes, nos considérations à la section 2 sur les désignateurs rigides et non rigides demeurent ici pertinentes. Pour nos propos dans cette section, accordons-nous simplement au sujet du fait que les théories de l'identité posent entre les états mentaux et les états physiques une corrélation d'identité stricte, de la même manière que « Eau = H_2O ». Il semble aussi évident selon les prémisses de la RM que le fonctionnalisme se dissocie de cette relation d'identité stricte, afin d'adopter une notion d'identité large. Donc, afin de respecter IDE,

il suffirait de montrer que N_1 et N_2 réalisent deux M identiques au sens large.

En gardant IDE en tête, concentrons-nous maintenant sur DIF. Il a été mentionné plus haut que le fonctionnalisme n'accorde pas d'importance à certaines différences contingentes, par exemple que les cerveaux des mammifères et des pieuvres diffèrent au niveau de leur localisation, du nombre, de la couleur, des atomes, etc. Ces propriétés ne devraient pas être considérées comme pertinentes. Pourquoi ? Pour imaginer ce propos, Shapiro donne l'exemple d'une série d'ouvre-bouteilles²⁹. Imaginons qu'il y ait trois ouvre-bouteilles placés devant nous. La seule différence entre eux est la couleur respective de chacun. Nous aurions donc un ouvre-bouteille rouge, un autre vert, puis finalement un jaune. Cette différence physique ne fait aucune différence sur ce qu'est, comme tel, un ouvre-bouteille. Les propriétés « rouge », « vert » et « jaune » ne sont ainsi pas pertinentes, puisque la fonction de l'ouvre-bouteille sera exactement la même peu importe la couleur de l'objet. Le fonctionnalisme en dira autant d'une série d'ouvre-bouteilles où l'un serait fait en bois, un autre en aluminium, puis un dernier en acier³⁰. Ces propriétés physiques ne font pas plus une différence quant à la fonction de l'artefact qui est d'ouvrir une bouteille de vin (quoique l'on puisse croire que les matériaux utilisés imposeront certaines contraintes physiques quant à la réalisation). Nous devons donc conclure que certaines propriétés physiques contingentes ne sont pas constitutives de l'ouvre-bouteille comme ouvre-bouteille puisqu'elles ne font aucune *différence* pour sa fonction d'ouvre-bouteille (non pas au sens de « différenciation », mais de « pertinence »). Selon Shapiro, pour que l'argument de la RM puisse en effet établir une relation DIF entre les réalisateurs (l'ouvre-bouteille comme objet matériel) et le réalisé (la fonction d'ouvrir les bouteilles), nous devons justement chercher les propriétés qui font une différence au niveau de la fonction instanciée³¹. Il s'ensuit que pour respecter DIF et IDE, nous devons rechercher une identité large des fonctions accomplies par des propriétés distinctes qui font une différence dans la réalisation physique.

Considérons maintenant le problème de la sorte : qu'arrive-t-il si deux objets font la même chose, mais exécutent cette fonction de

manière tout à fait différente? Soit leurs réalisateurs respectifs sont différents en ce qui a trait à leurs propriétés causalement pertinentes, soit au contraire ils sont isomorphes fonctionnellement. D'un côté, s'ils ne sont pas différents (c'est-à-dire s'ils ne respectent pas DIF), nous n'avons pas vraiment un cas de réalisabilité multiple. D'un autre côté, si les réalisateurs diffèrent quant à leurs propriétés pertinentes, nous pouvons nous demander en quoi ils ne seraient pas en fait de différents types fonctionnels. S'il s'avère que nous pouvons observer une différence au niveau des types fonctionnels, nous ne serions plus en présence de réalisabilité multiple. En résumé, la première option accepte un simple type de fonction, mais élimine la multiplicité de réalisateurs, alors que la deuxième option, toujours à être démontrée, bloquerait l'idée de réunir sous une même catégorie plusieurs réalisateurs. Lawrence A. Shapiro et Thomas W. Polger se sont attardés à la deuxième option problématique³². Suivant à notre tour le raisonnement de Shapiro et Polger, remarquons premièrement que lorsque nous regroupons des réalisateurs distincts, par exemple N_1 et N_2 , sous une même catégorie, nous les classons sous un même type en fonction du fait que N_1 et N_2 présentent certaines similarités que l'on peut *a posteriori* capturer sous des lois générales. Les cerveaux des mammifères (N_1) et ceux des pieuvres (N_2) diffèrent évidemment, mais certaines similarités sont observées permettant de conclure que tous deux sont aptes à réaliser une fonction F répondant aux spécifications C des structures neuronales N_1 et N_2 . Nous pouvons alors en déduire une loi générale qui permettrait d'inférer par la suite que lorsque nous sommes en présence de N_1 ou de N_2 , et que nous donnons un coup de marteau sur les têtes respectives de ces organismes, une douleur M devrait manifestement être instanciée en dépit des divergences au sein des structures neuronales.

Shapiro considère cependant que quand les réalisateurs sont significativement différents en ce qui a trait aux propriétés physiques pertinentes pour la fonction qu'elles exercent, les lois ou les généralisations tiennent difficilement la route³³. Prenons comme exemple que N_1 et N_2 sont tous deux aptes à instancier des propriétés visuelles. Nous pouvons ici dire qu'autant l'œil d'un primate que l'œil d'une pieuvre ont en commun d'opérer la même fonction,

c'est-à-dire voir. Nous sommes également bien en présence de deux structures différentes, puisque leurs yeux respectifs sont composés de pigments visuels différents, de photorécepteurs différents, de rétines différentes, etc. Ces propriétés doivent d'ailleurs faire une différence puisqu'elles réalisent des propriétés fonctionnelles (par exemple percevoir la lumière). Dès lors, le fonctionnalisme définirait la vision par sa fonction de percevoir la lumière, à partir de laquelle les processus mentaux (relativement à leurs rôles en tant qu'intermédiaire causal) analysent et interagissent avec l'environnement en fonction des « entrées » d'information et des « sorties » comportementales.

Nous pouvons concevoir, comme nous le disons dans le langage courant, que la pieuvre *voit*. Or, pouvons-nous à partir de ce constat général postuler une isomorphie fonctionnelle ? En effet, une étude plus fine des structures N_1 des mammifères (d'un primate) et N_2 des pieuvres, qui sont similaires, semble rapidement montrer des divergences au niveau des fonctions oculaires. Bien que tous deux possèdent des rétines, leurs méthodes respectives pour focaliser la lumière demeurent néanmoins différentes. Ces différences de structures des cellules photoréceptrices font en sorte que les primates (certaines espèces possédant une vision trichromatique, d'autres, dichromatique comme les daltoniens) peuvent percevoir les couleurs, là où la pieuvre ne le peut tout simplement pas. Les différences structurales au niveau des récepteurs visuels conduisent également à des différences temporelles de réaction. En fait, même la comparaison au sein d'espèces rapprochées pose de sérieuses difficultés à l'isomorphisme fonctionnel. Le règne animal en effet présente une variété de structures oculaires. Beaucoup de prédateurs ont des yeux binoculaires, permettant de mieux voir à distance. Les yeux de plusieurs proies, comme les souris et les lapins, sont plutôt placés sur les côtés, permettant alors de mieux repérer les dangers. D'autres, nyctalopes (les chats, certains rapaces), peuvent voir dans le noir, suggérant que la fonction de leurs yeux est celle de chasser dans le noir. En résumé, une analyse plus détaillée des fonctions réalisées par différentes structures confirme les propos de Shapiro pour qui les lois ou les généralisations tiennent mal la route suite à une analyse fonctionnelle entre espèces³⁴.

Dire dans le cas de la vision que N_1 et N_2 réalisent une isomorphie fonctionnelle que l'on nomme «voir» revient alors à une généralisation d'un niveau conceptuel plus élevé. Une analyse plus fine des fonctions oculaires dévoile plutôt des différences fonctionnelles occultées par le niveau descriptif auquel opère la généralisation fonctionnaliste. Au final, ces généralisations présentent des similarités isomorphes, mais non pas une relation d'identité comme la requiert la RM. Ce serait comme se méprendre au sujet d'un triangle possédant deux angles de 59° et un angle de 62° qui, bien qu'étant presque un triangle équilatéral, demeure par définition un triangle scalène. Pourtant, dans un sens plus général, il demeure qu'ils sont tous deux des triangles. Ce serait aussi se méprendre sur les deux éditions d'*Être et temps* que possèdent Kate et Delphine, l'une s'étant procuré l'édition de Gallimard, l'autre la traduction non officielle d'Emmanuel Martineau. Les deux éditions sont similaires, sans pour autant être identiques. En un sens plus général, ce sont bien deux exemplaires d'un même livre. Cependant, leurs différences de propriétés (la justesse de la traduction, les notes de bas de page, etc.) nous donnent bien deux œuvres différentes. Ce serait, au final, comparer des pommes et des oranges du simple fait que tous deux sont des fruits.

4. Le problème de l'héritage des pouvoirs causaux

Nous pouvons encore avancer un dernier argument, mais cette fois en ce qui concerne les pouvoirs causaux conférés au mental par ses réalisateurs physiques. Tel que précisé précédemment, le fonctionnalisme fait appel aux propriétés fonctionnelles afin d'expliquer les états mentaux. Nous arrivions alors à une formalisation selon laquelle

Pour toute propriété M , il y a une variété de propriétés physiques $P (P_1, P_2 \dots P_n)$, chacune répondant aux spécifications C du système S et qui instancient la propriété M .

M se définit alors pour le fonctionnalisme plutôt par ce qu'il *fait*, au lieu de ce qui le constitue. À partir du moment où l'on avance que P réalise M , il semble tout aussi logique de dire que M tire ses pouvoirs causaux de son réalisateur P . Tout comme un moteur ne peut avoir

de pouvoirs causaux au-delà des pouvoirs causaux rendus possibles par la structure physique accomplissant sa fonction spécifique de moteur, il s'ensuit selon la thèse fonctionnaliste qu'un état M ne peut avoir plus de pouvoirs causaux que ce que sa structure neuronale (et neurale) ne lui permet. Postuler le contraire impliquerait qu'il y a des pouvoirs causaux qui émergent hors du monde physique, ce qui reviendrait à contredire le principe de fermeture causale du physique. En vertu de ce qui relie les pouvoirs causaux du physique vers le mental, on constate également que M demeure en relation de dépendance à l'égard de P. Ainsi, non seulement le mental doit tirer ses pouvoirs causaux des propriétés physiques, la structure physique (fort probablement neurobiologique) est également essentielle à la réalisation du mental. Les thèses fonctionnalistes physicalistes non réductionnistes doivent alors admettre une dépendance des états mentaux par rapport aux états physico-chimiques du cerveau. Sans organisme capable d'instancier la douleur, celle-ci ne peut tout simplement pas avoir lieu.

Résumons dès lors l'essentiel de la problématique. Afin d'éviter le réductionnisme, il importe que les bases de réalisation de M soient physiquement différentes, et qu'ensuite M ait des pouvoirs causaux que ne possèdent pas ses réalisateurs physiques. En effet, il nous faut conserver l'efficacité causale des états mentaux, à défaut de quoi cela les rendrait superflus puisqu'ils seraient épiphénoménaux. En bref, il faut alors montrer que les structures réalisées présentent des différences causales par rapport aux structures réalisatrices (DIF), et ce, sans violer le principe de fermeture causale du physique. De plus, nous devons montrer que les M réalisés par deux structures neuronales différentes partagent des propriétés identiques (IDE). Le problème se résume alors à cette question : comment pouvons-nous expliquer la transaction causale entre le domaine physique et le domaine mental, tout en conservant la pertinence causale du mental alors qu'il est lié ontologiquement au physique, comme l'exige le physicalisme ? Il doit donc demeurer un *quelque chose de plus* aux phénomènes mentaux qui, causalement, les différencie de leurs bases de réalisation physiques (DIF), attendu qu'il s'agit bien du même état mental en dépit des bases de réalisation différentes (IDE). Si l'on

veut préserver les sciences particulières telles que la psychologie, il importe de montrer la pertinence causale des propriétés mentales. En effet, une science invoquant des phénomènes mentaux devrait nous procurer une explication de la pertinence causale de ces mêmes états mentaux. Comme le souligne Kim, un phénomène doit jouer un rôle explicatif dans une théorie qui démontrerait comment la présence ou l'absence de ce phénomène peut faire une différence causale sur le monde³⁵.

Nous avons déjà en grande partie répondu à la première partie du problème, soit le problème de différenciation et d'identité des fonctions causales de deux bases de réalisation présentant des différences au niveau des propriétés pertinentes. Tout en prenant en considération nos conclusions à ce propos, et en y ajoutant maintenant le problème des pouvoirs causaux, nous pouvons constater que le fonctionnalisme se retrouve de nouveau dans une impasse s'il veut éviter le réductionnisme. En premier lieu, si les pouvoirs causaux que M tire de P ne peuvent pas être causalement identiques à ceux qu'il tirerait de P_1 (ou $P_2 \dots P_n$), puisque les propriétés fonctionnelles des différentes bases de réalisation divergent à un niveau plus spécifique (telles étaient nos conclusions précédemment), P_1 doit spécifiquement réaliser M_1 . Si tel est le cas, M_1 survient seulement sur P_1 , M_2 sur P_2 , M_3 sur P_3 , etc. En deuxième lieu, il s'ensuit également que si les pouvoirs causaux de M sont hérités de P, nous pouvons en déduire le principe de l'héritage causal, formulé par Kim³⁶ :

Principe de l'héritage causal : Si une propriété mentale M est réalisée à t par une propriété physique P, alors les pouvoirs causaux de cette instanciation de M sont identiques à ceux de P.

Comme nous l'avons précédemment mentionné, affirmer que M a d'autres pouvoirs causaux que la structure physique qui l'instancie reviendrait à admettre des pouvoirs causaux hors du monde physique. Certes, plusieurs positions émergentistes soutiennent ce point, et critiquent même la conception de la causalité kimienne à partir de leur point de vue émergentiste³⁷. Aborder cette position dépasserait toutefois le cadre de cette discussion, centrée sur le fonctionnalisme. Accordons-nous pour dire qu'un fonctionnalisme

non réductionniste tel que formulé à partir des thèses de Putnam et de la RM doit quant à lui s'assurer que le monde physique demeure clos (principe de fermeture causale), tout en défendant l'irréductibilité et l'efficacité causale de M. Autant l'irréductibilité que l'efficacité causale de M deviennent problématiques pour le fonctionnalisme non réductionniste du moment où, en concordance avec le principe de l'héritage causal, les pouvoirs causaux de M s'avèrent en fait identiques à ceux de ses réalisateurs P.

Peut-être pourrions-nous plutôt avancer que les événements mentaux et physiques (et non les états) sont distincts. En effet, pour Kim, un événement est une exemplification par un objet (x) d'une propriété (P) à un temps t ($[x, P, t]$)³⁸. En suivant cette définition kimienne, deux événements, physique et mental, seraient identiques ssi $[x, P, t] = [y, M, t']$, c'est-à-dire si « $x = y$ », « $P = M$ » et « $t = t'$ ». Dire que, d'un côté, un individu exemplifie la propriété mentale M d'avoir mal à un temps t , et que, d'un autre côté, ce même individu exemplifie la propriété physique P d'avoir ses fibres C excitées à ce même temps t , semble effectivement nous amener à considérer deux événements distincts en vertu du fait que les propriétés exemplifiées diffèrent.

Ce que nous voulons donc déterminer, c'est si, pour un événement particulier, donc pour un même objet x , à un même temps t , P est réductible (ou non) à M. Le principe de l'héritage causal entre ici en ligne de compte. Si M et P partagent les mêmes pouvoirs causaux à un même temps t , il ne reste pas plus de raisons de croire que la propriété M diffère de P, même si nos descriptions des événements réfèrent à des propriétés différentes. Tout l'enjeu de cet argument réside encore une fois dans le fait que si nous admettons que les propriétés mentales ont des pouvoirs causaux distincts de leurs réalisateurs physiques, nous remettons du même coup le principe de fermeture causale en question, ce qu'un physicaliste ne serait sûrement pas enclin à faire.

Selon la définition même de propriété fonctionnelle, ce sont pourtant bien les propriétés physiques P (P_1, P_2, \dots, P_n) qui répondent aux spécifications du rôle causal que peut remplir une structure S. Comme nous l'avions déjà spécifié, l'état mental, en termes de

propriétés fonctionnelles, se définit par la fonction effective réalisée par P. Il apparaît alors juste d'affirmer qu'autant la fonction effective se base en fait sur la réalisation physique, autant le physique semble alors du même fait effectuer tout le travail causal. En effet, si, pour le fonctionnalisme, la notion d'état mental doit être comprise de la même manière qu'un concept d'artefact, par exemple comme le concept de moteur, mais qu'un moteur ne cause rien de plus que ce que lui permettent les propriétés physiques de sa structure, nous pouvons arriver à la conclusion que ce n'est pas le moteur qui cause la transformation d'une énergie en mouvement, mais bien les propriétés physiques qui répondent aux spécifications C de la configuration précise du moteur. Dans cette optique, nous pouvons aussi conclure que, soit M ne cause rien (et est donc inutile), soit M cause effectivement des événements, mais en vertu du fait que le mental, ontologiquement, est au final physique. Épistémologiquement, et de la même manière que l'on dit dans le langage courant qu'un moteur cause la motion d'un véhicule (alors que ce serait plutôt les propriétés physiques répondant aux spécifications causales du système moteur), les explications mentales peuvent être reléguées à des explications pragmatiques. Comme l'affirmaient les théoriciens de l'identité réductionniste à partir de la CMP, cela reviendrait à dire que les explications mentales ont recours à un langage différent, certes utile pour expliquer les processus cérébraux dans le langage quotidien de la psychologie populaire. Il n'en demeure pas moins qu'au niveau ontologique, la pertinence de M fait une fois de plus face aux menaces de l'épiphénoménisme, de l'exclusion causale et de la surdétermination. Nos conclusions concordent sur ce point avec celles de Kim pour qui les fonctionnalistes doivent, soit adopter une thèse épiphénoménaliste, soit admettre une réduction, dont j'ai montré qu'elle sera tout aussi multiple que le sont les bases de réalisation physique possibles. Au final, si tel est le cas, le fonctionnalisme non réductionniste apparaît donc inconséquent et ne semble pas être à la hauteur de ses prétentions, c'est-à-dire qu'il échoue ultimement à éviter la réduction des états mentaux aux états physiques.

Conclusion

En conclusion, l'adoption de la RM par le fonctionnalisme non réductionniste semble avoir de lourdes conséquences au niveau ontologique. L'un des avantages notoires de la théorie de l'identité est qu'elle pouvait nous procurer une simplicité ontologique. En effet, l'équation « $P = M$ » répond mieux aux exigences du rasoir d'Ockham, permettant des explications simples et efficaces des interactions du mental et du physique. En contrepartie, les conséquences tirées de la RM nous conduisent plutôt vers une multitude de réductions, alourdissant ainsi les possibilités de corrélation entre le mental et le physique. L'ontologie qui en découle doit à son tour comprendre une multitude d'entités. Cela n'implique pas qu'il n'y ait pas de réalisations de propriétés mentales à l'œuvre dans le cerveau, encore moins que nous devrions, face aux problèmes du fonctionnalisme non réductionniste, nécessairement adopter une position réductionniste. Les difficultés énoncées par Kim demeurent un défi de taille auquel tout postulat physicaliste devrait chercher à répondre. Le but de ce travail était justement de reconduire la RM et le fonctionnalisme vers ce défi kimien et ainsi de montrer qu'elle n'en est pas affranchie. De plus, si le fonctionnalisme non réductionniste ne peut pas répondre au principe de l'inférence vers la meilleure explication, c'est-à-dire procurer le meilleur *explanandum* aux états mentaux, il sera d'autant plus difficile à défendre qu'il semblerait alors perdre la bataille face au physicalisme réductionniste tant au niveau épistémologique qu'au niveau ontologique. Il en ressort dans tous les cas que l'argument de la RM n'est peut-être pas, dans sa forme fonctionnaliste classique du moins, adéquat pour éviter le réductionnisme tout en conservant l'efficacité causale des propriétés mentales.

-
1. Nous pouvons aussi parler de causalité du mental au physique et du mental au mental. Dans le premier cas, le mental causerait directement un événement physique (par exemple mon désir de boire un café m'a conduit dans la cuisine). Dans le deuxième cas, le mental aurait une influence directe sur d'autres événements mentaux (par exemple mon désir de boire une tasse de thé m'a fait penser à un ami au Japon).

2. Cf. John J. C. Smart, «Sensations and Brain Processes» dans *Philosophical Review*, vol. 68, 1959, pp. 141-156.
3. Cf. Jaegwon Kim, *Philosophie de l'esprit*, trad. David Michel-Pajus, Mathieu Mulcey et Charles Théret, Paris, Éditions Ithaque, 2008, p. 92. Cf. Jaegwon Kim, «Multiple Realization and the Metaphysics of Reduction» dans *Philosophy and Phenomenological Research*, vol. 52, n° 1 (mars 1992), p. 4.
4. Certains physicalistes, notamment les théoriciens du physicalisme éliminatif tels que Paul Churchland (1981) et Patricia Churchland (1986), argumentent en faveur de l'élimination de l'un des deux termes. En fait, si deux termes réfèrent exactement à la même idée ou à une même substance, on peut en effet mettre en doute l'utilité de s'y référer par deux expressions différentes, faute même, peut-être, de créer une confusion quant à la nature de cette substance. Cf. Patricia Churchland, *Neurophilosophy*, Cambridge, MA, MIT Press, 1986. Cf. Paul M. Churchland, «Eliminative Materialism and the Propositional Attitudes» dans *Journal of Philosophy*, vol. 78, n° 2 (février 1981), pp. 67-90.
5. Cf. Sophie Gibb, «The Causal Closure Principle» dans *The Philosophical Quarterly*, vol. 65, n° 261 (avril 2015), p. 628.
6. Jaegwon Kim, *Physicalism, or Something Near Enough*. Princeton, Princeton University Press, 2005, p. 43. Sophie Gibb, «Nonreductive Physicalism and the Problem of Strong Closure» dans *American Philosophical Quarterly*, vol. 49, n° 1, 2012, pp. 29-41.
7. Jerry Fodor, *A Theory of Content and Other Essays*, Cambridge, MIT Press, 1990, p. 156.
8. Lawrence A. Shapiro et Thomas W. Polger, «Identity, Variability, and Multiple Realization in the Special Sciences» dans Simone Gozzano et Christopher S. Hill, *New Perspectives on Type Identity*, Cambridge, Cambridge University Press, 2012, p. 266.
9. Cf. Hilary Putnam, «La nature des états mentaux», trad. Jean-Michel Roy dans D. Fisette et P. Poirier, *Philosophie de l'esprit, psychologie du sens commun et sciences de l'esprit*, Paris, Vrin, 2002, pp. 269-287.
10. Ned Block, «Introduction: What is Functionalism» dans Ned Block, *Readings in the Philosophy of Psychology*, Cambridge, Harvard University Press, pp. 171-184.
11. Jaegwon Kim, «Blocking Causal Drainage and Other Maintenance Chores with Mental Causation» dans *Philosophy and Phenomenological Research*, vol. 67, n° 1 (juillet 2003), p. 166.

12. Dans la suite de ce texte, «P» et «M» (sans souscrit) seront utilisés pour renvoyer à l'éventuelle propriété générique qu'exemplifient les différents $P_1, P_2 \dots P_n$ ou $M_1, M_2 \dots M_n$.
13. Putnam développe l'argument de la réalisabilité multiple dans une série d'articles datant des années 60. Nous référons principalement dans ce texte à l'article de 1967 «La nature des états mentaux». Cf. Hilary Putnam, *loc. cit.*, pp.269-287.
14. Putnam visait principalement la théorie de l'identité des types soutenue, entre autres, par Ullin T. Place, John C. Smart et Herbert Feigl. Cf. Ullin T. Place, «Is Consciousness a Brain Process?» dans *British Journal of Psychology*, vol. 47, 1956, pp. 44-50. Cf. John J. C. Smart, «Sensations and Brain Processes», *loc. cit.*, pp. 141-156. Cf. Herbert Feigl, «Le "mental" et le "physique"», trad. Christine Lafon et Bernard Andrieu, Paris, L'Harmattan, 2002, 304 pp.
15. John Bickle, «Multiple Realizability» dans *Stanford Encyclopedia of Philosophy*, [En ligne], <http://plato.stanford.edu/archives/spr2016/entries/multiple-realizability/> (Page consultée le 27 juin 2016)
16. Cf. Jaegwon Kim, *Philosophie de l'esprit*, *op. cit.*, p. 136.
17. *Ibid.*
18. Hilary Putnam, «Philosophy and Our Mental Life» dans Brian Beakley et Peter Ludlow, *The Philosophy of Mind: Classical Problems/Contemporary Issues*, Cambridge, MA, MIT Press, 1992, p. 91.
19. Saul A. Kripke, «Naming and Necessity» dans Donald Davidson et Gilbert Harman, *Semantics of Natural Language*, Dordrecht, D. Reidel Publishing Company, 1972, pp. 253-355.
20. Jaegwon Kim, *Philosophie de l'esprit*, *op. cit.*, p. 124.
21. *Ibid.*
22. David Papineau, «Causation is Macroscopic but Not Irreducible», dans S. C. Gibb, E. J. Lowe et R. D. Ingthorsson, *Mental Causation and Ontology*, Oxford, Oxford University Press, 2013, p. 131.
23. Le paradoxe de Fermi, d'après le nom du physicien italien Enrico Fermi, pose la question suivante: pourquoi n'avons-nous jamais trouvé de traces de vie extraterrestre, alors que pourtant notre soleil est beaucoup plus jeune que bien des soleils relativement proches de notre système solaire? Des civilisations extraterrestres, même éteintes, devraient en effet avoir laissé des traces (par exemple des ondes radio) captables de la Terre.
24. Nick Zangwill, «Variable Reduction Not Proven» dans *Philosophical Quarterly*, vol. 42, n° 167 (mars 1992), p. 218.

25. John Bickle, *Psychoneural Reduction: The New Wave*, Cambridge, MA, MIT Press, 1998.
26. Jitendra Sharma, Alessandra Angelucci et Mriganka Sur, «Induction of Visual Orientation Modules in Auditory Cortex» dans *Nature*, vol. 404 (avril 2000), pp. 841-847.
27. Lawrence A. Shapiro, «How to Test for Multiple Realization» dans *Philosophy of Science*, vol. 75, n° 5 (décembre 2008), p. 518.
28. Hilary Putnam, «La nature des états mentaux», *loc. cit.*, p. 281.
29. Lawrence A. Shapiro et Thomas W. Polger, «Identity, Variability, and Multiple Realization in the Special Sciences», *loc. cit.* p. 23. Cf. Lawrence A. Shapiro, *The Mind Incarnate*, Cambridge, MA, MIT Press, 2004.
30. *Id.*, «Multiple Realizations» dans *Journal of Philosophy*, vol. 97, n° 12 (décembre 2000), p. 644.
31. *Id.*, «How to Test for Multiple Realization», *loc. cit.*, p. 518.
32. Lawrence A. Shapiro et Thomas W. Polger, «Identity, Variability, and Multiple Realization in the Special Sciences», *loc. cit.*, pp. 264-286.
33. Lawrence, Shapiro, «Multiple Realizations», *loc. cit.*, p. 649.
34. *Ibid.*
35. Jaegwon Kim, «Blocking Causal Drainage and Other Maintenance Chores with Mental Causation», *loc. cit.*, p. 166.
36. Jaegwon Kim, «Multiple Realization and the Metaphysics of Reduction», *loc. cit.*, p. 18.
37. En résumé, la doctrine émergentiste avance que les agrégats de particules physiques, lorsqu'ils parviennent à un certain degré seuil de complexité structurelle, font émerger des propriétés nouvelles et authentiques. En somme, la devise des émergentistes peut se résumer à ce principe aristotélicien: «le tout est plus grand que la somme des parties». Cf. Jaegwon Kim, *Philosophie de l'esprit*, op.cit., pp. 326-334. Sur le problème de l'héritage causal et la position émergentiste, cf. Brian Garrett, «Héritabilité causale et propriétés émergentes» dans *Philosophiques*, vol. 27, n° 1 (printemps 2000), pp. 139-159.
38. Cf. Jaegwon Kim, «Causation, Nomic Subsumption, and the Concept of Event», dans *The Journal of Philosophy*, vol. 70, 1973, pp. 217-236.