



# **La théorie des tropes comme solution au problème de l'exclusion causale**

**Mémoire**

**Rodolphe Giorgis**

**Maîtrise en philosophie - avec mémoire**  
Maître ès arts (M.A.)

Québec, Canada

## Résumé

Dans ce mémoire il est question d'une position ontologique, la théorie des tropes, et de son application à une question centrale en philosophie de l'esprit : l'efficacité causale du mental. Nous présentons d'abord le problème de l'exclusion causale formulé par Jaegwon Kim afin de mettre en évidence les enjeux liés à cette question. L'argument a la forme d'un ensemble de thèses jugé inconsistant, et l'idée est de comprendre et de résoudre cette inconsistance, en rejetant l'une ou l'autre des thèses ou encore en clarifiant certains concepts mobilisés dans celles-ci. Une fois les thèses présentées et la contradiction expliquée, nous procédons à un tour d'horizon des solutions possibles, et nous nous arrêtons particulièrement sur la famille de solutions qui nous intéresse : le réductionnisme. Dans le second chapitre, nous présentons la théorie de tropes, position ontologique qui propose de désambigüiser le terme « propriété » afin de montrer qu'il est fréquemment utilisé pour désigner deux concepts différents. La propriété comme particulier d'une part, vue comme une entité unique, caractérisant une chose (la forme de cette pomme) ; et la propriété comme universel d'autre part, c'est-à-dire comme élément faisant le lien entre plusieurs choses semblant partager une même caractéristique (la couleur de ces deux livres, par exemple). La théorie des tropes affirme que le premier concept réfère aux tropes, des particuliers abstraits qui constitueraient l'ensemble de la réalité, et le second aux classes de ressemblance – des ensembles de tropes remplissant des rôles causaux identiques. Nous présentons donc ces deux concepts en détail. Enfin, dans le dernier chapitre, nous discutons les succès et les limites de cette approche. Nous commençons par montrer qu'il s'agit d'une position ontologique pertinente et légitime, puisqu'elle répond aux critiques métaphysiques qui lui sont adressées, et qu'elle est compatible avec une position cohérente en philosophie des sciences. Nous montrons ensuite comment David Robb et John Heil proposent d'utiliser la théorie des tropes pour formuler une position réductionniste en philosophie de l'esprit. Il s'agit de clarifier l'utilisation du terme « propriété » au fil de l'argument de l'exclusion, afin de résoudre la contradiction à laquelle il mène. Nous voyons ensuite que la théorie des tropes permet au réductionnisme d'échapper à l'une de ses plus vives critiques (l'argument de la réalisabilité multiple). Finalement, nous répondons à une critique de cette solution, qui suggère un retour du problème de l'exclusion au niveau des classes de ressemblance.

## Table des matières

Résumé .....	ii
Table des matières .....	iii
Remerciements.....	v
Introduction .....	1
Chapitre 1 : Le problème de l'exclusion causale.....	6
1.0 Introduction.....	6
1.1 Les thèses .....	8
1.1.1 Efficacité causale du mental.....	8
1.1.2 Physicalisme .....	12
1.1.3 Surdétermination causale .....	15
1.1.4 Fermeture du physique .....	17
1.2 Inconsistance des thèses : le problème de l'exclusion .....	18
1.3 Les solutions.....	23
1.3.1 Dualisme des substances .....	23
1.3.2 Épiphénoménisme .....	25
1.3.3 Fonctionnalisme : un cas de non-réductionnisme ? .....	27
1.3.4 Matérialisme éliminativiste .....	31
1.3.5 Réductionnisme .....	33
1.4 Conclusion.....	36
Chapitre 2 : La théorie des tropes.....	38
2.0 Introduction.....	38
2.1 La notion de propriété : une épine ontologique .....	38
2.1.1 Les propriétés : universaux ou particuliers ?.....	38
2.1.2 Les propriétés : dispositionnelles, catégoriques ou qualitatives ?.....	42
2.1.3 Le concept de substance .....	44
2.1.4 Propriétés et prédicats : la théorie de l'image ( <i>Picture Theory</i> ).....	46
2.2 Les niveaux de réalité.....	48
2.2.1 Niveaux de description et niveaux de réalité .....	48
2.2.2 Identité des types et identité des occurrences .....	50
2.2.3 La vision en strates rejetée .....	53
2.3 Classe de ressemblance et rôle fonctionnel .....	54

2.3.1 La technique Ramsey-Lewis et le projet d'analyse fonctionnelle de la notion de propriété.....	55
2.3.2 Théorie du rôle fonctionnel et ontologie .....	56
2.4 Conclusion.....	58
Chapitre 3 : Succès et limites de la théorie.....	60
3.0 Introduction.....	60
3.1 Trois critiques métaphysiques de la notion de tropes.....	60
3.2 Les tropes comme briques de construction : implications épistémologiques et métaphysiques .....	65
3.2.1 Réalisme et antiréalisme .....	65
3.2.2 Réalisme et théorie des tropes.....	68
3.3 La théorie des tropes, une solution réductionniste .....	71
3.4 La réalisabilité multiple : bête noire du réductionnisme? .....	75
3.4.1 L'argument de la réalisabilité multiple .....	75
3.4.2 Quelle réponse pour le réductionnisme ?.....	77
3.5 Le retour du problème de l'exclusion.....	78
3.6 Conclusion.....	82
Conclusion .....	83
Bibliographie.....	86

## Remerciements

Pour mener à terme ce mémoire de maîtrise, j'ai été accompagnée par une philosophe chevronnée et pertinente, qui est également une humaine généreuse, disponible et à l'écoute. Mes premiers remerciements vont donc à Renée Bilodeau, ma formidable directrice de maîtrise qui m'a fourni l'accompagnement et le support nécessaire, tout en sachant me laisser la liberté dont j'avais besoin. Je la remercie pour le temps qu'elle m'a accordé, pour ses conseils avisés et sa supervision éclairée. Je tiens également à lui témoigner toute ma gratitude pour les nombreuses opportunités qu'elle a mises sur mon chemin pendant mon parcours universitaire et pour sa confiance constamment renouvelée.

Je dois également exprimer mes plus sincères remerciements à certain.e.s collègues étudiant.e.s pour leurs conseils et leur écoute. Il serait impossible de nommer toutes celles et ceux qui ont contribué à mon développement intellectuel, mais j'aimerais mentionner en particulier Julien, qui m'a transmis sa passion pour la logique et son scepticisme à l'égard du libre arbitre, Romain qui a été un partenaire de correction exemplaire, David qui fut un colocataire fantastique, et Charles qui m'a écouté inlassablement parler de mes angoisses existentielles.

Je remercie également le Groupe de Recherche Interuniversitaire sur la Normativité (GRIN) et la Faculté de philosophie de l'Université Laval pour leur soutien financier.

La rédaction d'un mémoire de maîtrise est une entreprise qui occupe non seulement la personne qui le rédige, mais également tous ses proches. Je prends donc un peu du temps qui nous a manqué pour leur témoigner toute ma reconnaissance. Merci à mes parents, ainsi qu'à mes beaux-parents pour leur aide constante, leur écoute bienveillante, leurs encouragements intarissables et leur fierté rassurante. Merci également à mes ami.e.s, source inépuisable de divertissement, d'amitié et d'amour.

Enfin, je veux remercier celle sans qui rien de tout cela n'aurait été possible, sans qui j'aurais mille fois abandonné ce projet : mon épouse Maude, qui a su avoir la patience et la générosité de m'accompagner tout au long de ce processus de recherche et de rédaction.

## Introduction

Lorsque nous réfléchissons, délibérons, prenons des décisions et agissons, que se passe-t-il ? Deux intuitions à cet égard semblent coexister aussi bien dans le sens commun que dans la communauté scientifique. D'une part, nous attribuons ces actions et ces réflexions à des états mentaux : croyances, émotions, désirs, et intentions sont des éléments que nous considérons comme des causes et des conséquences de nos réflexions et des comportements qui les accompagnent. Des expressions comme « je crois que  $p$  », « vous désirez  $y$  », « elle a l'intention de  $z$  », sont omniprésentes dans nos discours et dans nos réflexions, et nous croyons qu'elles réfèrent à des entités réelles, agissant en nous : les états mentaux. Nous nous servons de ces expressions pour parler de nos problèmes de cœur, de nos relations amicales, de nos projets et de nos regrets. Nous attribuons ces états mentaux, les prêtons, les imputons parfois à nos semblables, et conversons ou délibérons avec eux des causes et des conséquences de nos actions ou de celles d'autrui en les expliquant par ces états mentaux.

D'autre part, ces réflexions et ces actions nous semblent être le produit de processus physiologiques : d'échanges électriques, physico-chimiques dans notre cerveau, dans notre système nerveux central et périphérique. En effet, nous voyons aisément que nos capacités à réfléchir et à agir sont directement liées à notre cerveau, et qu'elles sont affectées lorsque celui-ci est endommagé. De plus, cette intuition est confirmée par nos meilleures théories scientifiques, qui recensent et expliquent les corrélations entre nos réflexions, nos décisions et les actions qu'elles produisent d'une part, et notre activité cérébrale d'autre part (on peut penser à la célèbre expérience de Benjamin Libet<sup>1</sup>). Bien que de prime abord, la coexistence de ces deux intuitions puisse sembler anodine, un examen philosophique approfondi de leurs implications respectives montre qu'il est plus difficile qu'il n'y paraît de les concilier.

Pourtant, ni l'une ni l'autre ne semble extravagante ou particulièrement controversée. Nous avons fréquemment et avec succès recours à ces états mentaux pour décrire et expliquer nos comportements, ou encore pour prédire celui des autres et ajuster nos actions en fonction de ces prédictions. Par ailleurs, les neurosciences progressent et précisent toujours plus le lien entre nos comportements et l'activité physiologique de notre cerveau. On peut commencer à voir émerger la tension entre les deux intuitions : les états mentaux ne sont-ils que des états cérébraux, ou doit-on distinguer les deux ?

---

<sup>1</sup> Le neuroscientifique américain Benjamin Libet est connu pour avoir, à la suite d'expériences, émis l'hypothèse selon laquelle le cerveau humain initierait les actions de manière inconsciente. En effet, il aurait observé chez les sujets étudiés une activité cérébrale antérieure à la prise de décision des individus de réaliser une action volontaire. Ses travaux sont notamment consignés dans un article paru en 1999 et intitulé « Do We Have Free Will ».

Si on souhaite les distinguer, il nous faut expliquer comment des entités mentales jouent un rôle dans le monde physique. En son temps, Descartes proposait un dualisme des substances : il existerait deux types de substances distinctes et indépendantes, l'esprit et le corps.<sup>2</sup> Cette position est pour ainsi dire démodée de nos jours, car elle est problématique pour plusieurs raisons. Parmi elles, comment expliquer que l'esprit (ou le mental) agisse sur le corps (ou le physique), s'il en est détaché ? La position alternative est le monisme : il n'y aurait qu'une seule substance, soit le physique. Mais alors, qu'en est-il de l'intuition populaire, incarnée dans le langage par un lexique du mental omniprésent (« je désire  $x$  », « je crois que  $p$  », etc.), selon laquelle il existerait bien des états mentaux, qui ne seraient pas de simples échanges chimiques au niveau de nos synapses, ou des signaux électriques se propageant à travers nos neurones ? La tension dont nous parlons il y a quelques lignes semble s'accroître.

Si nous acceptons le monisme, et rejetons l'existence des états mentaux ou les réduisons simplement à des états physiques, comment garantir l'indépendance de la psychologie comme science, et son aspiration à élaborer des lois ? Celle-ci se base en effet sur des descriptions d'états mentaux pour développer son discours, et si ceux-ci peuvent être exprimés dans le langage des neurosciences, la psychologie risque de perdre ses compétences et son domaine d'expertise : l'étude de l'esprit, et la description, l'explication et la prédiction du comportement. Face à cette nécessité de rendre compte du statut du mental, les défenseurs du monisme ne s'entendent pas. Certains soutiennent qu'afin d'être cohérent, le monisme doit tout simplement réduire le mental au physique, d'autres avancent qu'il est possible de conserver une certaine indépendance du mental tout en acceptant l'existence d'un seul type de substance. Comme il est d'usage en philosophie, chaque position fait face à son lot de problèmes métaphysiques et épistémologiques.

Dans ce mémoire, nous allons nous intéresser à ce rapport entre mental et physique. En particulier, nous allons nous intéresser au physicalisme réductionniste et à une de ses versions : la théorie des tropes. Le physicalisme est sans doute le paradigme dominant en philosophie de l'esprit. Les physicalistes – pour le dire de façon très synthétique – croient que tout ce qui existe, tout ce qui constitue la réalité, est physique. Une des stratégies argumentatives les plus discutées par le courant analytique en philosophie de l'esprit dans les trente dernières années est celle du réductionniste Jaegwon Kim<sup>3</sup>, et consiste à présenter aux physicalistes – c'est-à-dire aux philosophes qui refusent le dualisme des substances – un choix entre le l'épiphénoménisme et le physicalisme réductionniste. Cet argument, c'est celui de l'exclusion causale. Il se targue de mettre en

---

<sup>2</sup> Il défend notamment cette position dans les *Méditations Métaphysiques*, parues en 1641.

<sup>3</sup> Cet auteur sera au cœur de nos discussions dans le premier chapitre de ce mémoire. Il a notamment publié deux ouvrages sur le sujet : *Philosophy of Mind* (une introduction aux questions majeures en philosophie de l'esprit, dont la troisième édition est parue en 2010), et *Physicalism, Or Something Near Enough* (une défense du réductionnisme parue en 2005). Il a également publié un grand nombre d'articles dans diverses revues et ouvrages collectifs. Nous aurons l'occasion de revenir sur plusieurs de ses textes dans les pages qui suivent.

évidence l'inconsistance d'un ensemble de thèses au sujet des états mentaux auxquelles nous semblons tenir, mais qui, considérées ensemble, font face à un problème de cohérence. Comme les physicalistes le soutiennent, tout ce qui cause des événements doit être physique, de sorte que le mental doit lui-même être physique – ou à tout le moins être très étroitement lié au physique. Dans un pareil cas, cependant, il semble compliqué de justifier l'idée que le mental est efficace causalement *en tant que mental*. Son efficacité causale semble exclue par celle des processus physiques desquels il dépend (ou avec lesquels il est identifié). Selon Kim, l'abandon de l'une ou l'autre de ces thèses est nécessaire afin de ne pas être empêtré dans une théorie contradictoire au sujet de l'esprit. Face à cette impasse, Kim affirme qu'il faut soit accepter le fait que les propriétés mentales ne sont pas efficaces causalement (l'épiphénoménisme), soit rejeter l'idée que le mental et le physique sont distincts, et accepter que le rôle causal traditionnellement attribué à certaines propriétés mentales est en fait rempli par les propriétés physiques auxquelles ces dernières peuvent être réduites, ou desquelles elles dépendent. La question à laquelle ce mémoire se propose de répondre est la suivante : dans quelle mesure la théorie des tropes apporte-t-elle une réponse réductionniste satisfaisante au problème de l'exclusion causale ?

Dans le premier chapitre, nous allons détailler l'argument de l'exclusion causale. Nous discuterons les thèses qui le composent et leur inconsistance. Nous présenterons ensuite différentes options qui s'offrent si l'on accepte l'argument : le dualisme, l'épiphénoménisme, le fonctionnalisme, le matérialisme éliminativiste, et, pour finir, le réductionnisme. La position réductionniste consiste à abandonner la thèse selon laquelle les propriétés mentales joueraient un rôle causal indépendant de leur base. Il serait en fait possible de les réduire à celle-ci. Pour expliquer cette réduction, plusieurs stratégies sont possibles. Les solutions qui misent sur la théorie de l'identité avancent que les propriétés mentales ne sont ni plus ni moins que des propriétés physiques. Toutefois, il y a un problème : un même type de phénomènes mentaux peut être réalisé de multiples façons par des processus physiques. L'exemple caractéristique est celui de la douleur. La douleur, un phénomène mental, peut être réalisée de différentes façons : par la stimulation d'un certain type de fibres chez les humains (les fibres C servent d'approximation dans cet exemple), mais par d'autres processus physiologiques chez d'autres espèces. Dans l'exemple classique de David Lewis sur la douleur des Martiens, ce sont des gonflements de certaines cavités sous les pieds d'une espèce extra-terrestre qui servent de base de réalisation physique à la douleur (Lewis, 1980). Cette expérience de pensée permettrait d'illustrer qu'il y a une distinction entre la douleur comme état mental d'une part, et ses différentes bases de réalisation physique d'autre part. Parce qu'on peut imaginer qu'un extra-terrestre qui ne possède pas de cerveau peut quand même ressentir de la douleur, il faudrait distinguer les états mentaux des états physiques : on ne pourrait se contenter de réduire les premiers au second. Cette objection à la position réductionniste est aussi utilisée comme argument en faveur du fonctionnalisme, nous la discuterons donc car il nous faudra défendre notre version de la position réductionniste face à elle.



Cette version, nous la détaillerons dans le second chapitre. Il s'agit de la théorie des tropes. Elle a été développée par D.C. Williams au milieu du XX<sup>e</sup> siècle (Williams, 1953a, 1953b) et reprise par de nombreux théoriciens et théoriciennes de l'esprit. Nous nous référerons un peu à Williams, mais nous nous concentrerons sur des textes de David Robb et de John Heil, d'abord, et aussi d'Ann Whittle, car ces philosophes s'intéressent particulièrement à l'application de la théorie des tropes à la question du rapport entre mental et physique.<sup>4</sup> En effet, si nous l'abordons presque exclusivement sous cet angle dans ce mémoire, la théorie des tropes est une théorie ontologique dont les différentes thèses ont des implications qui dépassent les limites de la philosophie de l'esprit. De façon très générale, une des thèses principales de cette théorie affirme que la réalité est constituée de tropes : des particuliers abstraits. Appliquée à l'argument de l'exclusion, la théorie permet de désambiguïser la notion de propriété – qui contiendrait en fait deux idées : ce qui appartient en propre à une entité et lui donne sa nature particulière (ce qui caractérise l'objet ou l'évènement en question), et l'élément commun que partagent plusieurs objets distincts (la propriété d'être de couleur bleue partagée par ma tasse et par ma bouteille d'eau). La première idée réfère à un trope, c'est l'idée d'une propriété particulière, possédée par un objet. La seconde réfère à ce qui est parfois vu comme un universel : nous verrons que cette notion vient avec son lot de difficulté et que nous pouvons lui préférer la notion de type ou de classe de référence. Dans un premier temps, nous présenterons donc la notion de propriété et ses différents enjeux. Puis, nous nous intéresserons à la question des niveaux de réalité et aux différentes théories de l'identité, afin de voir à laquelle la théorie des tropes peut se rapporter. Pour finir, nous préciserons la notion de classe de ressemblance en montrant qu'il est possible de concilier une ontologie de tropes avec une conception fonctionnaliste de la notion de propriété au sens universel.

Enfin, dans le troisième chapitre, nous discuterons des succès et des limites de cette théorie. Nous aborderons d'abord quelques problèmes auxquels elle est confrontée, afin de justifier sa pertinence comme position ontologique. Nous aborderons une série de trois critiques métaphysiques, adressées directement à la thèse ontologique de la théorie des tropes, et nous verrons en quoi elle peut répondre à celles-ci. Nous verrons également en quoi elle est compatible avec une position cohérente en philosophie des sciences. Ensuite, nous montrerons en quoi elle peut être utile dans le cadre du problème corps-esprit : en apportant une solution réductionniste à l'argument de l'exclusion causale. Grâce à elle, on peut en effet conserver l'unité d'une notion comme la douleur en termes de classe de ressemblance, mais distinguer entre plusieurs tropes différents tombant sous cette même classe de ressemblance (en fonction de l'espèce par exemple). Par

---

<sup>4</sup> Heil, Robb et Whittle sont trois philosophes dont les travaux portent notamment sur la théorie des tropes et sur son application en philosophie de l'esprit. Tous trois ont publié des textes au XXI<sup>e</sup> siècle et il nous semble donc plus pertinent de discuter leurs positions dans le détail plutôt que celle de Williams, qui n'est plus vraiment discutée aujourd'hui même si elle continue d'exercer une certaine influence – notamment sur nos trois philosophes. Parmi leurs travaux, on peut citer (Heil 2003, 2012), (Robb 2013) et (Whittle 2007, 2008).

ailleurs, un même trope peut aussi tomber sous la coupe de plusieurs classes de ressemblance différentes. Si l'on reprend l'exemple de la douleur, et que l'on considère la douleur de J.R.R. Tolkien s'étant brulé en allumant sa pipe, on a affaire à un trope précis : la douleur de Tolkien à ce moment précis. Toutefois, ce trope tombe sous la classe de ressemblance des fibres humaines stimulées ainsi que sous la classe de ressemblance des états mentaux de douleur. Le Martien qui se blesse les tentacules ressent une douleur précise à un moment donné : il s'agit à nouveau d'un trope précis. Ce trope tombera bien sous la classe de ressemblance de la douleur, mais il ne tombera pas sous la classe de ressemblance de la stimulation des fibres C. À la place, il tombera sous la classe de ressemblance correspondant au processus physique réalisant la douleur chez son espèce. Les propriétés mentales peuvent donc bien être réduites, identifiées à des éléments physiques en cela que les deux ne forment en fait qu'un seul et même trope, et les classes de ressemblance maintiennent à la fois une certaine indépendance du mental par rapport au physique et une unicité des états mentaux à travers leurs multiples réalisations. Pour finir, nous aborderons une objection soulevée par Paul Noordhof quant à la capacité de celle-ci à répondre au problème de l'exclusion (Noordhof, 1998). Pour lui, en effet, le problème de l'exclusion revient au niveau des tropes, et la question qui pouvait être source d'ambiguïté lorsqu'on parlait de propriété peut maintenant être posée clairement à propos des tropes : en vertu de quoi un trope est-il efficace causalement ? Nous verrons que pour Robb, cette critique est basée sur une incompréhension de la théorie et de son application au problème soulevé par Kim.

# Chapitre 1 : Le problème de l'exclusion causale

## 1.0 Introduction

Nos croyances, nos désirs, nos intentions se causent les uns les autres et causent nos actions. Ces croyances, ces désirs et ces intentions sont liés à notre cerveau. Ces deux intuitions largement répandues au sujet du mental semblent avoir de bonnes raisons de l'être : la première car nous sommes passés maîtres dans l'art de prédire les actions des autres ou encore d'expliquer les nôtres en ayant recours au champ lexical du mental. Vu le succès de ces prédictions et explications, il semble donc raisonnable de penser qu'une réalité sous-jacente doit correspondre à celles-ci. La popularité de la seconde intuition, quant à elle, peut être expliquée par le fait que nous nous rendons vite compte par l'expérience qu'un coup à la tête peut affecter notre expérience consciente et notre capacité à raisonner et à agir, voire les rendre nulles. Toutefois, la conciliation de ces deux intuitions pose plus de problèmes qu'on ne pourrait le croire. Bien que nous ayons de la facilité à communiquer nos croyances et nos désirs sous la forme de phrases telles que « je veux un café » ou « je crois qu'Angela Davis est l'autrice de *Are Prisons Obsolete ?* », il apparaît difficile, aujourd'hui encore, de saisir avec précision le lien entre ces états mentaux et notre cerveau, bien que nous soyons convaincus qu'il y en a un.

En philosophie de l'esprit, les débats portant sur les rapports entre les états mentaux et le cerveau tournent le plus souvent autour d'une position introduite dans les années 1930 par Rudolph Carnap : le physicalisme. Très grossièrement, adhérer à celle-ci revient à croire que tout ce qui existe est physique. Nous aurons l'occasion de la clarifier au fil de la section, mais il est important de garder en tête que c'est autour de cette position qu'ont lieu la plupart des débats dans ce champ de la tradition analytique. L'alternative, le dualisme, est à notre époque peu défendue : il s'agirait d'affirmer que les états mentaux ne sont pas physiques, ne sont pas des états de notre cerveau, mais sont d'une substance autre, une substance mentale détachée du domaine physique, mais liée à lui et interagissant causalement avec lui. Le philosophe français René Descartes est notamment connu pour avoir défendu une position de ce type. Bien qu'il en existe des versions contemporaines, ces thèses restent toutefois très minoritaires parmi la communauté philosophique.

Suffit-il alors d'affirmer que tout est physique pour en avoir terminé avec la question de l'efficacité causale du mental ? Pouvons-nous nous contenter d'affirmer que nos états mentaux *sont* et *ne sont que* des états cérébraux, des interactions physico-chimiques entre nos neurones à travers nos synapses, dont les neurosciences auront tôt fait de nous révéler le fonctionnement détaillé ? Cela a le mérite d'expliquer également la première intuition : les états mentaux peuvent se causer entre eux et causer nos actions puisqu'ils appartiennent eux-mêmes au domaine physique. Les choses ne sont cependant pas si claires. Au sein même du physicalisme, deux familles de théories se font face. Les réductionnistes défendent globalement

l'approche que nous venons de décrire. Ils pensent que l'on peut *réduire* les états mentaux à des états cérébraux. Les non-réductionnistes, eux, maintiennent qu'il est possible de conserver une certaine indépendance du mental par rapport au physique, des états mentaux par rapport aux états cérébraux, tout en restant dans un cadre physicaliste. Nous verrons dans ce chapitre comment s'articulent ces différentes positions, en nous intéressant à un argument utilisé par les réductionnistes pour montrer que le non-réductionnisme est une position incohérente : l'argument de l'exclusion causale.

L'argument de l'exclusion est énormément discuté dans la littérature, mais c'est au philosophe Jaegwon Kim que l'on doit sa formulation originale. Ce dernier tente de montrer que le non-réductionnisme implique d'abandonner l'efficacité causale des états mentaux au profit de celle des entités physiques sur lesquelles ses défenseurs n'ont pas le choix de les faire dépendre – sans quoi ceux-ci basculeraient dans le dualisme. Nous allons revenir en détail sur cet argument, sa forme et ses conséquences. Pour ce faire, nous allons présenter différentes intuitions qui sont à l'origine de la position physicaliste et de sa grande popularité. Les quatre intuitions qui vont nous intéresser sont discutées par les philosophes de l'esprit, et certaines sont ensuite formulées en des thèses, plus systématiques et plus concises, avec un lexique plus précis. Développées et organisées, ces thèses permettent d'obtenir une position philosophique cohérente au sujet de l'esprit et de son rapport causal au monde. La première intuition est celle selon laquelle nos actions auraient des causes mentales. La deuxième nous pousse à croire que le mental dépendrait, au moins dans une certaine mesure, de notre cerveau. Ce sont les deux intuitions dont nous venons de parler. La troisième intuition affirme que tout effet physique doit avoir une cause physique, et la quatrième qu'un effet n'a généralement<sup>5</sup> qu'une seule cause.

L'argument de l'exclusion tente de montrer que l'ensemble des thèses que sous-tendent ces intuitions est incohérent – ou à tout le moins, semble l'être. La combinaison de ces intuitions commence plutôt bien. Considérons une action accomplie par un agent (comme frapper une balle lors d'un match de tennis, par exemple). La première intuition nous porte à croire que cette action aurait une cause mentale, le désir du joueur d'envoyer la balle à un certain endroit du court dans le but de remporter l'échange. Or, selon la seconde intuition, ce désir dépendrait au moins dans une certaine mesure d'états du cerveau du joueur. Cette idée est confortée par la troisième intuition, qui nous pousse à chercher une cause physique (disons, un état cérébral précis se caractérisant par le déclenchement de certaines synapses ou encore par l'excitation de certains neurones) à l'action de notre agent, puisque celle-ci est elle-même physique (le corps de l'agent se déplace) et a une influence sur d'autres objets (et potentiellement sur d'autres agents). Toutefois, la dernière

---

<sup>5</sup> Cette intuition reste neutre sur l'existence d'authentiques cas de surdétermination, mais elle affirme qu'ils ne sont pas systématiques. Nous revenons sur cet enjeu dans la section 1.1.3.

intuition souligne qu'il faut éviter de postuler l'existence de plusieurs causes pour le même effet : il nous faut faire un choix entre l'état mental et l'état cérébral, et ce même si le premier dépend du second. Pour Kim, c'est là qu'intervient le dilemme du non-réductionniste : il lui faut accepter que l'état mental ne fasse finalement qu'un avec l'état cérébral duquel il dépend, faute de quoi le rôle causal est rempli par le second, et le premier s'en voit privé, et il faut donc renoncer à l'efficacité causale du mental.

La solution qui nous intéresse dans ce mémoire appartient à la famille des positions réductionnistes. Elle consiste à désambiguïser la formulation des thèses qui sous-tendent chacune des intuitions afin d'empêcher la contradiction que nous venons de présenter brièvement en proposant une identité entre la cause physique et la cause mentale. Dans ce chapitre, nous allons parler en détail de chacune des intuitions et formuler les thèses correspondantes le plus précisément possible (1.1), puis nous allons expliciter la contradiction qu'elles entraînent (1.2), et enfin les différentes solutions qui se présentent (1.3).

## 1.1 Les thèses

### 1.1.1 Efficacité causale du mental

Le caractère répandu de la première intuition est mis en évidence par le fait que nous référons constamment aux causes et aux conséquences de nos actions en parlant d'états mentaux : « j'ai fait du café parce que j'en avais *envie* », « je fais attention à ce que je mange parce que *je crois* que cela va me permettre d'être en meilleure santé », « Fumer cette cigarette *m'a changé les idées* ». Dans ces exemples, il semble que les causes (avoir envie de café, croire que l'alimentation joue un rôle dans la santé) ou que les conséquences (se changer les idées) de nos actions sont de nature mentale. À tout le moins, les prédicats dont nous nous servons pour référer à ces causes et à ces conséquences appartiennent au champ lexical du mental. Cette idée selon laquelle les états mentaux peuvent causer des événements se retrouve aussi dans des disciplines médicales ou théoriques, comme la psychiatrie ou la psychologie : les praticiens attribuent tel ou tel problème à une peur, une mauvaise gestion de désirs ou des croyances injustifiées, etc. L'intuition selon laquelle les actions des agents ont des causes mentales conduit les théoriciens de l'esprit à formuler et à discuter une thèse, celle de l'efficacité causale du mental.

Pour formuler cette thèse, il nous faut introduire plus précisément la notion d'état mental. Afin de ne pas entrer tout de suite dans les débats entre les différentes positions que nous allons discuter plus loin, nous allons utiliser une définition la plus large et la plus générale possible. Nous entendons par état mental tout état d'un agent qui se rapporte à son esprit. Par « esprit », nous n'entendons pas une dimension mystique détachée du corps, nous désignons simplement les états liés à la réflexion, au ressenti et au traitement des données sensorielles. Ainsi, nous pouvons notamment distinguer les perceptions des sensations, les premières étant les processus mentaux d'interprétation et de traitement des secondes. Outre les perceptions,

l'ensemble des états mentaux contient également ces états auxquels nous référons à l'aide du vocabulaire du mental : croyance, peur, envie, désir, certitude, sont de bons exemples. Une catégorie d'états mentaux particulièrement discutée en philosophie de l'esprit contient les attitudes propositionnelles. Il s'agit d'états mentaux ayant un *contenu propositionnel*, c'est-à-dire d'états mentaux dont le contenu est une proposition. Par exemple, si Sarah croit que la terre est ronde, nous dirons qu'elle a un état mental dont le contenu est « la terre est ronde ». Cet état mental consiste à croire que la proposition « la terre est ronde » est vraie.

À l'aide de la notion d'état mental, on peut formuler la première thèse ainsi :

[Efficacité causale du mental, *première version*] : Certains états mentaux causent des évènements physiques.

Toutefois, cette version de la thèse (bien que reflétant adéquatement l'intuition présente dans le discours populaire et en psychologie) est trop imprécise. Il est question d'une relation causale, mais quels éléments sont engagés dans cette relation ? Quelle est la nature de cette relation ? La notion de causalité est abondamment discutée dans plusieurs champs d'étude et ne fait bien sûr pas l'objet d'un consensus. Puisqu'il serait trop ambitieux de l'examiner sous toutes ses coutures et de présenter chacun des problèmes qu'elle soulève, nous allons nous contenter de présenter la conception dont nous nous servirons ici, et d'en défendre brièvement la pertinence.

Deux éléments sont essentiels pour définir pour toute conception de la causalité : les *relata* et la relation. La définition de la relation permet de distinguer une relation causale d'un autre type de relation (d'une relation de corrélation par exemple)<sup>6</sup> ; le choix des *relata*, lui, permet d'avoir une ontologie cohérente : l'idée est de décider quel type d'entités sont impliquées dans une interaction causale. Parlons-nous d'évènements qui en causent d'autres ? Parlons-nous de propriétés conférant des pouvoirs causaux, est-il question de faits, ou encore d'états de choses (*state of affairs*) ? Nous allons discuter dans un premier temps les différentes conceptions de la relation de causalité, puis nous ferons quelques remarques sur l'aspect ontologique – la question des *relata*.

Il existe plusieurs approches de la relation de causalité. On peut les diviser en trois catégories : les approches interventionnistes, régularistes, et contrefactuelles. On retrouve déjà les deux dernières dans les travaux de David Hume, qui écrit à la fin de la section VII de son *Traité sur l'Entendement Humain* : « we may define a cause to be *an object, followed by another, and where all the objects, similar to the first, are followed*

---

<sup>6</sup> Pour une discussion des relations de dépendance non causale, voir le second chapitre du livre de Kim, *Supervenience and Mind*, intitulé « Noncausal connections » (Kim 1974, p. 22-32).

*by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed* » (Hume 1748, p. 56). Dans ce court passage, le philosophe empiriste exprime déjà deux des approches principales de la notion de causalité. La première est l'approche régulariste et elle a été systématisée par John Stuart Mill à partir des travaux de Hume. Pour les régularistes, deux caractéristiques suffisent à qualifier une relation de relation causale : le fait qu'un effet *suive* sa cause, et le fait qu'un effet *suive toujours* sa cause, que les deux aillent systématiquement de pair. La première caractéristique rend compte du lien que l'on fait entre la notion de causalité et celle de temporalité : il semblerait profondément contre-intuitif d'affirmer qu'un effet puisse être antérieur à sa cause. La seconde, quant à elle, la distingue de simples associations fortuites d'évènements se succédant dans le temps mais n'ayant pas de lien causal.

Dans la deuxième partie du passage, Hume donne une autre définition de la notion, qui se rapproche davantage de ce que l'on appelle aujourd'hui les approches contrefactuelles de la causalité. Un énoncé contrefactuel est un énoncé qui postule une conséquence de l'absence d'un évènement sur l'état du monde. Un énoncé de ce type a la forme « si l'évènement *e* ne s'était pas produit, alors *B* ». Notons que « *B* » désigne ici un énoncé qui peut affirmer l'absence ou l'occurrence d'un autre évènement postérieur au premier. On peut donc imaginer que « *B* » se lise « l'évènement *f* se serait produit », ou « l'évènement *f* ne se serait pas produit ». Dans l'extrait considéré plus haut, Hume semble définir la relation de causalité à l'aide d'une analyse contrefactuelle. Cette idée a longtemps semblé être une simple reformulation de son approche régulariste, exprimée juste avant et détaillée ailleurs dans son œuvre. Toutefois, elle sera reprise plus tard par plusieurs philosophes, comme David Lewis, pour définir la notion de causalité uniquement à l'aune de l'analyse contrefactuelle (Lewis 1973).

La troisième famille d'approches de la notion de causalité part du constat de l'échec de l'analyse contrefactuelle à en fournir une théorie complète et cohérente. La plupart des approches interventionnistes sont des extensions de l'analyse contrefactuelle, à laquelle sont ajoutées des considérations qui ne se limitent pas au domaine conceptuel, mais qui postulent une influence de l'expérience, des observations et du raisonnement causal lui-même sur la notion de relation causale. Judea Pearl, par exemple, propose une analyse de la notion de causalité s'inspirant de l'approche contrefactuelle et de la théorie des probabilités<sup>7</sup> (avec des considérations comme « une cause augmente les chances que son effet se produise »), mais qui fait également une place aux effets de l'intervention d'observateurs sur les raisonnements causaux

---

<sup>7</sup> Certaines approches de la notion de causalité se basent uniquement sur une analyse probabiliste pour en rendre compte, mais elles sont minoritaires. Le plus souvent, les probabilités jouent un rôle dans une analyse de la notion (comme chez Pearl) mais n'en constituent pas l'intégralité. Pearl soulève d'ailleurs dans ses travaux plusieurs problèmes liés au fait d'analyser la causalité uniquement à l'aune des probabilités.

(Pearl 2018).<sup>8</sup> Nous n'entrerons pas davantage dans les détails de chaque approche, et nous nous servirons par défaut de l'approche contrefactuelle dans le reste de ce mémoire, puisqu'elle est utilisée par la plupart des philosophes que nous allons discuter.

Outre ces débats sur la meilleure manière d'analyser la relation de causalité, un débat oppose plusieurs conceptions ontologiques de la causalité elle-même, à travers notamment la question des *relata*, c'est-à-dire des éléments impliqués dans les relations de causalité. Lorsque nous parlons de causalité, nous allons parler d'une relation entre deux évènements. Puisque nous avons retenu l'approche contrefactuelle, nous pouvons retenir la formule suivante : un évènement C cause un autre évènement E si et seulement si, si C ne s'était pas produit, E ne se serait pas produit non plus. Cette approche est défendue par David Lewis, qui la présente notamment dans l'article « Causation », paru en 1973 et considéré comme un article phare dans la littérature moderne sur la causalité. Aucune position sur la causalité ne fait consensus, bien entendu, mais pour les besoins de notre discussion, nous allons donc considérer les causes et les effets comme des évènements liés entre eux par des relations que l'on peut exprimer à l'aide de l'analyse contrefactuelle.

Mais poussons le questionnement ontologique un peu plus loin : en vertu de quoi un évènement peut-il garantir qu'un autre se produise ? Autrement dit, *comment* un évènement en cause-t-il un autre ? Il existe bien sûr plusieurs positions sur la question, mais nous allons retenir celle de Kim, puisque nous allons nous intéresser à son problème de l'exclusion. Il définit un évènement comme un triplet : l'instanciation d'une *propriété* par un *objet* (il est parfois question de « substance », concept sur lequel nous revenons dans les chapitres suivants) à un *moment* du temps. Il écrit dans *Supervenience and Mind* : « According to this account, each individual event has three unique constituents: a substance (the “constitutive object” of the event), a property it exemplifies (the “constitutive property” or “generic event”), and a time. An event is a complex of these three, and I have used the notation [x, P, t], or variants thereof, as a canonical notation for events in general » (Kim 1976, p. 35). Dans le second chapitre de ce mémoire, nous nous intéresserons plus longuement à la notion de propriété, mais pour le moment, contentons-nous de considérer les propriétés comme porteuses de ce qu'on peut appeler des « pouvoirs causaux ». Un évènement en cause un autre parce qu'il est l'instanciation de certaines propriétés, et ce faisant, ces dernières lui confèrent un certain nombre de pouvoirs causaux en vertu desquels il garantit que son effet se produit. L'intérêt de cette précision est de

---

<sup>8</sup> Voir à ce sujet l'ouvrage de Pearl, *Causality : Models, Reasoning and Inference*. Ses conclusions sur l'importance de faire une place à l'intervention des observateurs dans les raisonnements causaux partent de constats sur une contradiction manifeste entre nos intuitions au sujet de la causalité et la façon dont on la formule traditionnellement d'une part, et des résultats statistiques curieux comme le paradoxe de Simpson d'autre part. Le sixième chapitre du livre est particulièrement pertinent à ce sujet (Pearl 2018, p. 173).



déplacer le débat sur l'aspect mental et physique au niveau des propriétés, plutôt que de parler d'évènements mentaux et d'évènements physiques.

C'est donc au niveau des propriétés que va se jouer la question de l'exclusion : un évènement (un évènement cérébral par exemple) en cause-t-il un autre (l'action d'un agent) en vertu du fait qu'ils sont l'instanciation d'une propriété physique (un courant électrique parcourant le cerveau par exemple, ou encore un échange chimique entre des synapses) ou en vertu du fait qu'ils sont l'instanciation d'une propriété mentale (l'état mental correspondant) ? L'intuition selon laquelle nos désirs et nos croyances sont effectivement des causes et des effets semble pointer vers la deuxième solution. Grâce à ces notions, il est donc possible de formuler la thèse de l'efficacité causale du mental un peu plus précisément :

[Efficacité causale du mental] : Certains évènements en causent d'autres en vertu du fait qu'ils sont l'instanciation d'une propriété mentale.

Cette version de la thèse, outre le fait qu'elle est plus précise que la précédente, a aussi l'avantage de couvrir à la fois les cas de causalité entre évènements mentaux (un désir en causant un autre) et les cas de causalité entre des évènements non cérébraux et des évènements cérébraux (une croyance ou un désir causant un mouvement), au sens où elle porte sur tous les évènements et déplace l'enjeu de l'efficacité causale au niveau des propriétés.

### **1.1.2 Physicalisme**

Une autre intuition importante est celle selon laquelle le mental dépend, au moins dans une certaine mesure, de ce qui se passe dans notre cerveau. Dans nos interactions quotidiennes, la chose est très claire : beaucoup de lieux communs concernant l'intelligence, le jugement, la réflexion, etc. comportent les mots « tête » ou « cerveau ». De plus, il est très fréquent de voir dans les travaux en neurosciences des conjectures cherchant à associer des évènements cérébraux à des états mentaux. Une position très discutée en philosophie de l'esprit traduit en partie cette intuition : le physicalisme. Version moderne du matérialisme<sup>9</sup>, cette thèse peut être formulée de plusieurs façons, certaines plus fortes et plus engageantes que d'autres. L'idée générale est que tout ce qui existe est constitué d'éléments étudiés par les sciences physiques : atomes, électrons, particules élémentaires constitueraient la totalité de la réalité. Une des versions les plus

---

<sup>9</sup> Le matérialisme est en quelque sorte l'ancêtre du physicalisme. Il s'agit, très succinctement, de la position selon laquelle tout ce qui existe est constitué de matière. Bien que certains philosophes se servent encore du terme « matérialisme », on lui préfère plus souvent qu'autrement le terme « physicalisme », car celui-ci se rapporte de façon plus précise aux éléments dont les sciences physiques font l'étude. De plus, le matérialisme, dans sa forme classique, exclut certaines réalités physiques qui ne sont pas de la matière bien qu'ils soient clairement physiques – c'est le cas des champs magnétiques.

faibles du physicalisme (car peu engageante et peu controversée) est le physicalisme dit survenant, ou *supervenience physicalism*. Avant de le présenter, arrêtons-nous sur la notion de survenance.

Celle-ci peut être facilement illustrée : considérons le tableau d'un.e artiste peintre. Celui-ci a un certain nombre de propriétés macroscopiques qui peuvent être observées par des individus d'un simple coup d'œil : il peut être de forme carré, rectangulaire, ovale, ou autre ; il peut représenter un personnage, un objet, un paysage, etc. Pourtant, ce tableau n'est qu'un agencement particulier de peinture sur une toile. On peut dire que les propriétés représentationnelles macroscopiques du tableau sont *survenantes* sur les agencements de peinture sur la toile. Le tableau ne pourrait pas exhiber d'autres propriétés macroscopiques sans que l'agencement de la peinture ne soit différent. De façon générale, on dit qu'un ensemble ( $S^*$ ) de propriétés survient sur un autre ensemble ( $S$ ) de propriétés si et seulement s'il ne peut y avoir de différence dans  $S^*$  sans qu'il y en ait dans  $S$ .

Revenons au physicalisme. L'intérêt du physicalisme survenant est qu'il nous permet de faire au mental une place qui semble cohérente avec les intuitions que nous avons à son sujet : il existe et semble avoir une influence sur le monde extérieur, mais il n'est pas indépendant ontologiquement du domaine physique. De plus, ce type de physicalisme n'est pas trop engageant et contrairement à d'autres versions de la thèse, il est assez largement accepté.<sup>10</sup> Ainsi, il est compatible avec presque toutes les solutions au problème de l'exclusion dont il est question dans la littérature. Sans nous intéresser plus que nécessaire aux différentes implications du physicalisme, il est important de mentionner qu'y adhérer<sup>11</sup> nous oblige au moins à postuler une certaine dépendance du mental sur le physique. Ce que nous pouvons faire à l'aide de la notion de survenance. Ajoutons donc à notre ensemble de thèses :

[Survenance du mental] : Les propriétés mentales surviennent sur des propriétés physiques.

Si l'on suit la définition de la notion de survenance donnée plus haut, cette thèse implique donc bien une *dépendance ontologique* des propriétés mentales sur les propriétés physiques. Plus précisément, deux conséquences immédiates découlent de la thèse. Tout d'abord, l'instanciation d'une propriété mentale dépend

---

<sup>10</sup> Les fonctionnalistes préfèrent parler de physicalisme de la réalisation (*realization physicalism*). La notion de réalisation est plus forte que celle de survenance. En conséquence, s'il existe une relation de réalisation entre les états mentaux et les états physiques, il existe aussi une relation de survenance, mais non l'inverse. Nous n'entrerons pas dans les nuances entre les différentes positions. Sauf quand nous discuterons la thèse de la réalisabilité multiple, nous laisserons la notion de réalisation de côté et utiliserons plutôt celle de survenance.

<sup>11</sup> Nous verrons un peu plus loin que rejeter le physicalisme est la stratégie des dualistes, qui postulent l'existence de plusieurs substances distinctes (pour les dualistes, une réplique physique du monde ne serait pas une réplique complète du monde). Puisque nous ne considérerons pas cette possibilité en détail, nous allons tenir le physicalisme survenant, tel que défini plus haut, comme un axiome implicite régissant notre réflexion, dont [Survenance du mental] est une conséquence, et nous ne le traiterons donc pas directement dans notre ensemble de thèses.

de l'instanciation d'une propriété physique sur laquelle celle-ci survient. C'est ce qui fait de cette position une position physicaliste : une propriété mentale ne peut exister indépendamment d'une base physique. La deuxième conséquence de [Survenance du mental] est que si une propriété physique P est une base de survenance pour une propriété mentale M, alors tout système qui instancie P instancie également M au même moment. Cette conséquence illustre le caractère nécessaire de la relation de survenance, tout en rappelant le sens de la dépendance entre physique et mental : pour qu'il y ait une différence au niveau du mental entre deux états d'un système donné, il *faut* qu'il y ait une différence au niveau de la configuration physique de ce système. Si deux états d'un même système présentent une différence au niveau du mental (s'il manque une propriété mentale dans l'un des deux, par exemple), cela implique qu'ils présentent également une différence au niveau du physique. L'inverse n'est pas vrai : deux états du même système peuvent différer au niveau de leur structure physique sans qu'on observe de changements au niveau des propriétés mentales qu'ils instancient. Il suffit pour cela que les propriétés physiques qui les distinguent n'instancient pas de propriétés mentales.

Cette thèse rend justice à une intuition importante dont nous avons parlé : le mental dépend, au moins dans une certaine mesure, du cerveau. Toutefois, il ne semble pas souhaitable de réduire tout à fait le mental au physique : nous voulons conserver une certaine indépendance entre l'un et l'autre. La première raison pour cette volonté d'indépendance concerne le champ lexical du mental, dont il a déjà été question plus haut : son efficacité et son omniprésence dans les interactions sociales ainsi que dans plusieurs champs pratiques et théoriques donnent à penser que les éléments qu'il désigne existent bel et bien et doivent être conservés, si possible, lors de réflexions ontologiques. On peut alors formuler une thèse d'irréductibilité du mental au physique, en suivant ici encore le modèle de Kim dans son livre *Physicalism, or Something Near Enough*.

[Irréductibilité] : Les propriétés mentales ne sont ni réductibles ni identiques à des propriétés physiques.

Pour expliquer ces deux caractéristiques, Kim écrit : « [...] physically irreducible properties remain outside the physical domain – that is, if anything is physically reduced, it must be identical with some physical item [...] If Xs are reduced to Ys, then Xs are nothing “over and above” brain processes » (Kim 2005, p. 34). On tente donc, avec ces deux nouvelles thèses, de faire une place aux deux intuitions de départ dans notre réflexion. Comme nous le verrons un peu plus bas, il n'est toutefois pas clair qu'il soit possible de les conserver toutes les deux.

### 1.1.3 Surdétermination causale

Une intuition au sujet de la causalité mène à l'affirmation d'une thèse supplémentaire qui entre en jeu dans l'argument de l'exclusion : il s'agit de la surdétermination causale. Lorsqu'il est question de causalité, nous avons en tête une certaine relation qui lie deux éléments. Cette relation peut être représentée, nous l'avons vu, de multiples façons : causalité régulariste, causalité contrefactuelle, etc. Toutefois, une intuition commune à son sujet est qu'un effet est en quelque sorte *généré* par sa cause. En cela, on a tendance à considérer cette dernière comme unique ; comme l'unique responsable (causalement) de l'avènement de son effet. On ne semble pas vouloir qu'il existe plusieurs causes suffisantes pour le même effet. Si l'on affirme qu'un événement  $C_1$  est la cause d'un autre événement  $E$ , il semble contre-intuitif de dire que l'on pourrait tout aussi bien affirmer qu'un autre événement  $C_2$  est la cause du même événement  $E$ . Quelle serait alors la spécificité de  $C_1$  par rapport à  $C_2$  ? Qu'est-ce qui pourrait justifier de le qualifier de cause de  $E$  ?

Cette intuition est représentée plus ou moins fortement dans les différentes conceptions de la causalité, mais elle est particulièrement claire dans le cas de l'analyse contrefactuelle que nous avons définie plus tôt. Rappelons que celle-ci, lorsqu'appliquée à la notion de causalité, exprime la relation entre une cause et son effet de la façon suivante :  $C$  cause  $E$  si et seulement si, si  $C$  ne s'était pas produit,  $E$  ne se serait pas produit non plus. Dès lors, on semble tenir au principe selon lequel un effet n'aurait qu'une seule cause, puisqu'on exprime littéralement la relation causale comme dépendant du fait que l'absence de la cause résulte en l'absence de l'effet. Or s'il existe plus d'une cause pour un même effet, en l'absence de celle-ci l'effet pourrait tout de même se produire, en autant qu'au moins une de ses causes ait lieu.

Dans son argument, Kim fait alors appel au principe d'exclusion, qui refuse que la surdétermination soit monnaie courante. Une autre thèse à prendre en compte dans notre ensemble peut donc se lire :

[Exclusion] : Un événement ne peut généralement pas avoir plus d'une cause suffisante.

Plusieurs remarques sont de circonstance. Tout d'abord, le terme « généralement » est employé pour rendre compte des cas authentiques de surdétermination. Un exemple abondamment discuté dans la littérature est celui de deux tireurs d'élite ( $A$  et  $B$ ) tirant en même temps une balle dans le cœur de leur victime. La mort de la victime semble être surdéterminée par les deux tirs. On voit aisément que ce cas pose problème au modèle contrefactuel : le contrefactuel « Si  $A$  n'avait pas tiré la victime ne serait pas morte » est faux, de même que « Si  $B$  n'avait pas tiré la victime ne serait pas morte ». On semble pourtant tenir à ce qu'au moins un des deux tirs soit la cause de la mort. Quelle est alors la cause de la mort de la victime ? Nous ne pouvons ni ne voulons entrer dans des discussions concernant l'existence, la pertinence ou la fréquence de ces cas limites, et nous nous contentons donc de postuler qu'il s'agit, s'ils existent, d'exceptions. En revanche, puisque toute

propriété mentale doit être réalisée par une propriété physique, il semble que tous les événements qui ont une cause mentale seraient systématiquement surdéterminés. Ils auraient toujours à la fois une cause mentale et une cause physique. Cela en ferait des cas uniques dans la nature, où les cas de surdétermination (s'ils existent) sont l'exception, non la règle. Puisque nous ne souhaitons pas analyser les chaînes causales contenant des causes mentales comme des cas à part du reste des relations causales, il semble donc pertinent de faire une place au principe d'exclusion dans notre ensemble de thèses, afin de ne pas faire de ces cas d'évènements causés par des états mentaux des évènements à part.

De plus, même si la surdétermination était réelle, qu'elle n'était pas juste anecdotique mais fréquente, répandue ; si, donc, tous les évènements ayant des causes mentales étaient surdéterminés, cela mènerait à une autre forme de contradiction. Bien qu'elle semble de prime abord apporter une conciliation entre l'idée de l'efficacité causale du mental et le physicalisme (en permettant au mental d'être efficace causalement en vertu de sa dépendance à un élément physique sans rendre celui-ci impotent causalement), la surdétermination mène à une conclusion indésirable. En effet, si l'on accepte – pour les besoins de la démonstration – la surdétermination d'un évènement par une propriété mentale et une propriété physique, on doit accepter qu'en l'absence de la propriété physique, la propriété mentale fera le travail toute seule. Cela semble contredire le physicalisme, et reconduire à une forme de dualisme : une situation dans laquelle il serait difficile d'expliquer *comment* quelque chose de non physique pourrait jouer un rôle causal dans le domaine du physique. Kim se demande, dans le passage suivant, si la surdétermination est une option sérieuse à considérer :

Could it be that the mental cause and the physical cause are each an *independent sufficient* cause of the physical effect? The suggestion then is that the physical effect is *overdetermined*. So if the physical cause hadn't occurred, the mental cause by itself would have caused the effect. This picture is again absurd: from what we know about the physiology of limb movement, we must believe that if the pain sensation causes my hand to withdraw, the causal chain from the pain to the limb motion must somehow make use of the causal chain from an appropriate central neural event to the muscle contraction; it makes no sense to think that there might be an independent, perhaps telekinetic, causal path from the pain to the limb movement (Kim 1993b, p. 281).

Il semble ainsi qu'accepter la surdétermination mène à une impasse. Pour éviter ce genre de situation que Kim juge absurde, il paraît donc approprié d'ajouter le principe d'exclusion à notre ensemble de thèses.

Kim mentionne quelques lignes plus loin qu'une autre raison de rejeter la surdétermination est intimement liée au problème que l'on vient de soulever : celle-ci est incompatible avec une autre thèse à laquelle on semble avoir de bonnes raisons de vouloir adhérer. Il écrit : « moreover, the overdetermination idea violates the causal closure principle as well » (Kim 1993b, p. 281). Tournons-nous maintenant vers cette notion de fermeture causale, ou fermeture du domaine physique.

### 1.1.4 Fermeture du physique

Nous avons vu plus haut que les intuitions concernant le physicalisme, et les thèses [Survenance du mental] et [Irréductibilité] qui en découlent ne nous engagent pas à grand-chose concernant la façon dont le mental peut agir sur le physique. Par [Efficacité causale du mental], nous avons déjà postulé l'existence d'une telle influence. Toutefois, une autre intuition répandue au sujet de la causalité est à prendre en compte, et joue un rôle important dans l'argument de l'exclusion : celle selon laquelle le physique n'est causé que par du physique. L'idée est que nous sommes capables, lorsque nous tentons d'expliquer l'histoire causale d'un événement observé dans le monde, de le faire en ne référant qu'à des causes physiques. Parmi les raisons qui poussent les philosophes à adhérer à une telle thèse, on peut encore une fois citer la volonté d'éviter l'écueil dualiste : pourquoi postuler des éléments appartenant à un autre niveau de réalité pour retracer l'histoire causale d'un événement, alors que nous évoluons dans un paradigme majoritairement physicaliste ? Il semble plus simple d'avoir recours aux éléments étudiés et mis à notre disposition par les sciences naturelles. Nous pouvons formuler la thèse correspondant à cette intuition de la façon suivante :

[Complétude du physique] : Si un événement a une cause à  $t$ , il a une cause physique à  $t$ .

La fermeture du physique est une thèse plus controversée que [Physicalisme], et tout comme cette dernière, des versions plus ou moins fortes sont discutées dans la littérature. Kim en formule deux ; une faible ([Complétude du physique]), et une forte : « [Strong Closure] Any cause of a physical event is itself a physical event – that is, no nonphysical event can be a cause of a physical event » (Kim 2003, p. 162). La seconde est plus forte car elle est d'emblée incompatible avec [Efficacité causale du mental]. Elle rend donc toute autre discussion futile : il faut choisir entre les deux sans quoi l'on tombe dans une contradiction immédiate. [Complétude du physique], elle, est compatible avec [Efficacité causale du mental]. Kim suggère que son utilisation est préférable dans le cadre de l'argument de l'exclusion. La raison de ce choix est que la version faible est moins engageante et donc plus apte à établir une base de réflexion pertinente pour discuter du statut du mental. Nous allons revenir sur cet aspect de base de réflexion, qui est une des grandes forces de l'argument, dans la section suivante. Dans cet extrait, Kim explique qu'il vaut mieux adopter une version faible de la thèse, car cela permet aux non-réductionnistes de ne pas quitter la discussion trop tôt.

Adopting [Strong closure] as a premise is like starting your argument with mind-body causation already ruled out, at least for nonreductivists; with [Strong closure] as your starting point, there isn't very much more distance you can go or need to go [...] It is better, that is to say, to distribute the burden of proving a conclusion among a set of relatively weak premises than to place that burden on fewer but individually stronger premises. The latter strategy is apt to provoke the complaint that the argument begs the question and that it serves no useful purpose. I think we learn something about the issues and desiderata involved and their interplay when we run the supervenience argument with [Closure] rather than [Strong closure] (Kim 2003, p.163).

Par ailleurs, [Complétude du physique] a l'avantage d'avoir la forme d'une implication matérielle : elle est formulée à l'aide d'un conditionnel. Ainsi, elle est vraie (de façon vide) même s'il n'existe pas de relation de causalité liant les événements, et ne met donc pas notre ensemble de thèses en péril.

Voyons à présent comment ces différentes thèses interagissent, et s'il est possible de rendre intelligible notre conception du mental et de donner à nos deux intuitions de départ la place qui leur est due.

## 1.2 Inconsistance des thèses : le problème de l'exclusion

L'argument de l'exclusion – ou à tout le moins, la version que je présente ici<sup>12</sup> – est dû à Kim, qui l'a formulé dans plusieurs articles et ouvrages dans lesquels il défend sa position concernant la nature des propriétés mentales : le réductionnisme. Il s'agit d'un raisonnement par l'absurde : l'argument a la forme d'un ensemble de thèses qui s'avère être inconsistent. Il faut donc abandonner au moins l'une des thèses de l'ensemble afin de ne pas générer de contradiction et ainsi qu'il reste consistant. La force de l'argument réside dans sa simplicité conceptuelle et dans sa capacité à poser efficacement les bases du débat sur la nature des propriétés mentales. Avant de présenter la structure générale puis le détail de l'argument, il convient de discuter ces deux avantages afin de comprendre l'intérêt qu'il suscite et la richesse de la littérature à son sujet.

Commençons par la simplicité de l'argument. L'argument est simple à la fois dans sa forme et dans son contenu. Comme nous l'avons vu dans les dernières sections, il est constitué d'un ensemble de thèses qui se basent sur des *intuitions* au sujet du mental, de la causalité et de la structure physique du monde. Ces intuitions sont très répandues et elles ne semblent pas farfelues : la thèse physicaliste, par exemple, semble plus en phase avec l'état actuel de nos connaissances et de nos théories scientifiques que ne le serait un dualisme des substances qui postulerait des propriétés mentales détachées du domaine physique<sup>13</sup>. Ainsi, prendre ces intuitions comme base pour construire l'argument place la réflexion qui en découle au carrefour de la psychologie et des sciences naturelles (neurobiologie et physique en particulier), ce qui a pour effet de la rendre pertinente. En effet, l'intuition selon laquelle le mental est responsable – causalement responsable – de nos actions est à la base de la psychologie et de la pratique psychiatrique. Au risque de nous répéter, le champ lexical du mental est omniprésent dans ces deux domaines et il ne semble donc pas controversé

---

<sup>12</sup> On en retrouve en effet une première version dans l'article « The Conceivability of Mechanism » de Norman Malcolm, publié en 1968. Toutefois, c'est la version de Kim qui a suscité un intérêt majeur en philosophie de l'esprit.

<sup>13</sup> Cela n'est bien sûr pas une raison suffisante de rejeter a priori toute position dualiste. Toutefois, puisqu'il est parfois reproché aux philosophes de ne pas prendre en compte le sens commun, il est pertinent de préciser que les intuitions qui se cachent en arrière des thèses dont nous discutons sont des intuitions répandues.

d'affirmer que ceux-ci postulent l'efficacité causale des éléments auxquels ils réfèrent.<sup>14</sup> Par ailleurs, l'utilisation d'intuitions concernant la structure physique du monde, desquelles on a entre autres tiré [Survenance du mental], fait de la discussion générée par l'argument une discussion qui prend en compte l'état actuel de la connaissance humaine, en entérinant le paradigme dominant dans lequel se trouve la communauté scientifique : le physicalisme. Il nous semble important, pour conserver une certaine crédibilité, de prendre en compte l'état actuel des connaissances scientifiques ; ce qui ne veut pas dire se limiter à leurs conclusions, ni même adhérer au réalisme scientifique – la position selon laquelle il est raisonnable de croire que les éléments décrits par les sciences constituent effectivement la réalité – comme position métaphysique. Après tout, comme nous le verrons dans le second chapitre, la théorie des tropes remet en question cette position, et adhérer à celle-ci semble même nous approcher davantage de l'antiréalisme<sup>15</sup>.

Le deuxième avantage de l'argument est qu'il offre plusieurs solutions. Le but de Kim lorsqu'il le formule n'est pas de défendre une position en particulier, mais de proposer un cadre à l'intérieur duquel il est possible de discuter de la nature des propriétés mentales, et de la manière dont elles jouent un rôle causal dans la production d'événements. Son propre point de départ est d'ailleurs d'essayer de rendre cohérente la position non réductionniste<sup>16</sup>. L'argument de l'exclusion serait donc davantage un point de départ pour la réflexion qu'une étape décisive de celle-ci. La structure de l'argument est d'ailleurs propice à l'utilisation de l'argument comme d'une base de discussion. Considérons les grandes thèses présentées à la section précédente comme des axiomes : des positions que nous tenons pour vraies, et à partir desquelles nous allons faire une série d'affirmations (que nous nommons théorèmes) en suivant des règles de déduction aussi rigoureuses que possible. Lorsqu'un problème émerge, qu'il s'agisse d'une contradiction directe entre certains théorèmes ou entre un théorème et un axiome, cela force à se questionner sur l'origine de la contradiction : le raisonnement était-il bien valide ? Si c'est le cas, alors il faut renoncer à l'un ou l'autre des axiomes, puisqu'ils nous mènent à une contradiction. La discussion peut alors porter sur la question de savoir quel(s) axiome(s) il convient d'abandonner et sur les raisons qui motivent cette décision.

---

<sup>14</sup> L'alternative serait d'imaginer que la psychologie et la psychiatrie ne seraient que des pansements appliqués à des problèmes qu'on ne pourrait réellement traiter qu'avec des neurosciences et une neurochirurgie de pointe, ou « arrivées à maturité », pour reprendre l'expression des éliminativistes.

<sup>15</sup> La position antiréaliste est d'ailleurs bien décrite par Kim dans *Philosophy of Mind* : « If you think of scientific theories in general, or theories over some specific domain, merely as predictive instruments that enable us to infer or calculate further observations from the given ones, you will not attach any existential significance to its theoretical posits--like unobservable microparticles of theoretical physics and their (often quite strange) properties--and regard them only as calculational aids in deriving predictions. A position like this is called "instrumentalism" or "antirealism" about scientific theory » (Kim 2010, p. 149)

<sup>16</sup> Voir à ce sujet son article « Making Sense of Emergence », et la section sur le physicalisme non réductionniste dans son livre *Physicalism Or Something Near Enough* (p. 33).



La structure de l'argument est donc simple : à partir d'une liste restreinte d'affirmations (nos axiomes, tels que présentés dans la section précédente), il s'agit d'exprimer la manière dont les propriétés mentales s'y prennent pour remplir le cahier des charges que ces axiomes leur imposent : être causalement efficaces, dépendantes du physique, et non réductibles à celui-ci. On peut saisir facilement l'idée générale de l'argument à l'aide de la figure ci-dessous.

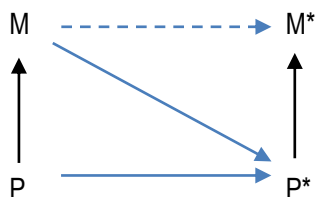


Figure 1

Par [Efficacité causale du mental], on postule l'existence de M, qui cause l'instanciation d'une autre propriété mentale, M\*. [Survénance du mental] garantit que M et M\* aient des bases de réalisation physique : respectivement P et P\*. De plus, par [Complétude du physique], P\* doit avoir une cause *physique* complète et suffisante : on peut alors supposer que cette cause est P. Cela nous permet aussi de rendre cohérente la façon dont M cause M\* : selon la formule de Kim, « M caused M\* by causing its supervenience base P\* » (Kim 2005, p. 40). C'est ce qui explique que la flèche représentant le lien causal entre M et M\* soit en pointillés. Cependant, par [Non-surdétermination], M et P ne peuvent pas causer P\* tous les deux. Il semble donc que les thèses soient inconsistantes : elles ne peuvent pas être toutes vraies en même temps.

Avant de nous pencher sur les tensions entre nos différentes thèses, regardons ce raisonnement un peu plus en détail. Tout d'abord, on postule l'existence d'une propriété mentale M qui est efficace causalement, celle-ci va donc causer M\*<sup>17</sup>. Cette étape nous est donnée par l'axiome [Efficacité causale du mental], à l'aide duquel on postule un événement qui en cause un autre en vertu du fait que le premier est l'instanciation d'une propriété mentale par un objet à un moment du temps. Par [Physicalisme], il faut également postuler P et P\*, comme base de survénance respective de M et de M\*. Cela nous donne les deux premières étapes du raisonnement :

<sup>17</sup> Tel que mentionné plus haut, c'est en fait l'instanciation de la propriété M par un objet à un moment du temps (un événement) qui va causer l'instanciation de M\* par un objet à un moment du temps (un autre événement). Nous allégeons ici le texte en supprimant les références constantes aux événements et au temps.

- (1) M cause M\*. (Par [Efficacité causale du mental])
- (2) M\* survient sur P\*, sa base de survenance physique. (par [Survenance du mental])

Ici, une tension émerge entre (1) et (2). Si M cause M\*, et que M\* survient sur P\*, il semble difficile de trancher sur ce qui est responsable de M\* : est-ce M (sa cause) ou P\* (sa base de survenance) ? Ce questionnement préliminaire nous amène à nous demander s'il est nécessaire de présenter M comme la cause de M\*, ou s'il suffit d'expliquer P\* pour expliquer M\*, puisque cette dernière survient sur P\*<sup>18</sup>. Cette tension est en fait à la base de la réflexion de Kim sur l'exclusion causale. Il s'agit de la tension entre la causalité, une relation horizontale (diachronique) et la détermination, une relation verticale (synchronique). L'idée générale est qu'en admettant que [Survenance du mental] est vraie, la présence de P\* (la base de survenance de M\*) à un instant *t* est *suffisante* pour garantir l'occurrence de M\* à *t*. L'histoire causale de M à M\* ne semble pas pertinente. Comme le note Kim : « The tension between vertical determination and horizontal causation, or the former's threat to preempt and void the latter, has been, at least for me, at the heart of the worries about mental causation » (Kim 2005, p. 38). Heureusement, une solution assez simple permet de résoudre – au moins pour le moment – cette tension entre (1) et (2). Puisque l'occurrence de M\* est inévitable si P\* se produit, il semble que M ne soit pas pertinent pour expliquer M\*, à moins que M soit requis pour causer P\*. C'est précisément la prochaine étape du raisonnement :

- (3) M cause M\* *en causant* sa base de survenance P\*.

À partir de cette affirmation seule, nous pouvons déduire :

- (4) M est une cause de P\*.

La prochaine étape fait appel à l'intuition selon laquelle un événement physique, s'il a une cause, doit avoir une cause physique. Nous avons un événement physique, P\*, et nous commençons à tracer son histoire causale. Par [Complétude du physique], puisque P\* a une cause à un instant *t* – cause qui est postulée par (4) – alors il doit avoir une cause *physique* à *t*. Or, pour continuer à être cohérent, nous pouvons postuler que cette cause est la base de survenance de M. Il y a plusieurs raisons pour cela. Tout d'abord, M et P se produisent en même temps. De plus, cela semble aller de pair avec (3). Nous avons donc :

---

<sup>18</sup> Cela peut donner lieu à une scission de la discussion du problème de l'exclusion causale : il peut être traité de façon *verticale* (la question porte alors sur la relation de survenance, et nous amène à nous demander si celle-ci est assez forte pour rendre caduc l'écueil de la surdétermination), ou de façon *horizontale* (le débat porte alors sur la relation de causalité, et l'enjeu est de savoir si cette dernière peut accommoder les différentes exigences que nous avons au sujet du mental). Pour une discussion de ce point de vue sur le problème de l'exclusion, voir Donaldson (2021).

(5)  $P^*$  a une cause physique  $P$ .

L'histoire pourrait se finir ici, si nous décidions d'identifier  $M$  à  $P$ . Toutefois, comme nous l'avons vu au début de ce chapitre, il est préférable de ne pas s'aventurer sur ce chemin : nous voulons conserver l'indépendance du mental vis-à-vis du physique. C'est ce qui a motivé l'axiome d'[Irréductibilité], et c'est avec lui que l'argument se poursuit.

(6)  $M \neq P$ .

Par (4), (5) et (6), nous avons :

(7)  $P^*$  a deux causes.

Par [Exclusion], et en supposant que nous ne soyons pas en présence d'un cas de surdétermination, nous avons alors :

(8) Il faut éliminer  $M$  ou  $P$  comme cause de  $P^*$ .

L'argument touche alors à sa fin, car [Complétude du physique] revient à la charge pour nous inciter à éliminer  $M$ . Nous avons alors :

(9)  $M$  est abandonnée comme cause de  $P^*$ .

Le rôle causal de  $M$  est alors *exclu* par celui de  $P$ , sa base de survenance. Cela va à l'encontre de notre hypothèse de départ, selon laquelle  $M$  est la cause de  $M^*$ . Puisque ces notions ne semblent pas fonctionner toutes ensemble, il semble alors intéressant de discuter les issues possibles : que doit-on conserver ? Que doit-on modifier ? Que doit-on abandonner tout à fait ? Le problème de l'exclusion prend tout son sens en tant que cadre de réflexion.

La solution apparente est d'abandonner une des thèses dont on se sert dans la démonstration. L'abandon de [Non-surdétermination] soulève les questions présentées succinctement plus haut en autorisant un effet à avoir plus d'une cause suffisante : pour Kim, qui qualifie la solution d'absurde, elle n'est pas envisageable. Si l'efficacité causale du mental est abandonnée, le coût est l'épiphénoménisme : le postulat d'une propriété,  $M$ , qui existerait sans le moindre pouvoir causal, sans qu'elle ne change quoi que ce soit à la production de l'effet (puisque  $P$  serait toujours là pour le causer). Nous n'entrerons pas dans les détails tout de suite, mais la chose semble contre-intuitive : les différentes conceptions de la causalité ont en commun l'idée d'une certaine nécessité de l'occurrence d'une cause pour la production d'un effet ; le fait qu'il soit possible

que M n'existe pas et que son effet se produise quand même semble mener à la conclusion, *contra-hypothesis*, que la première n'est pas la cause du second.

L'abandon de [Complétude du physique] est la stratégie préférée par de nombreux non-réductionnistes : elle permet de faire une place au rôle causal unique, et non déterminé par P, de M.

Les réductionnistes, comme Kim, préfèrent abandonner [Efficacité causale du mental] ; ou plutôt désambiguïser cette dernière. Les propriétés mentales seraient réductibles à leur base physique, et seraient donc efficaces causalement *en tant que* propriétés physiques, respectant ainsi [Complétude du physique]. Nous allons nous pencher sur certaines des solutions les plus classiques à ce problème.

### 1.3 Les solutions

Les solutions au problème de l'exclusion proposées dans la littérature sont extrêmement nombreuses. Il est toutefois possible de les classer en plusieurs catégories, car beaucoup d'entre elles partagent la même idée directrice. Tel que mentionné plus haut, la solution apparente semble être d'abandonner un ou plusieurs des axiomes de départ afin d'éviter la contradiction. C'est ainsi que nous allons classer les solutions abordées ici<sup>19</sup>. Nous allons d'abord considérer tour à tour les conséquences de l'abandon de l'une ou l'autre des thèses avant de présenter la catégorie de solutions dont fait partie celle qui nous intéresse dans ce mémoire.

#### 1.3.1 Dualisme des substances

La première famille de solutions tient à tout prix à la prémisse [Irréductibilité], elle préfère la conserver et abandonner plutôt n'importe quelle autre thèse. La position la plus connue parmi celles-ci est sans doute le dualisme des substances. Avant d'évaluer une version plus contemporaine du dualisme, il est intéressant d'en considérer l'origine. Bien qu'on puisse en identifier des versions plus anciennes (dans les travaux des théoriciens médiévaux ou des Grecs), c'est chez Descartes qu'on trouve les bases théoriques de cette famille de positions – à tout le moins dans sa forme actuelle. Pour ce dernier, il existe deux substances séparées, irréductibles l'une à l'autre, et pouvant être distinguées sur la base de deux caractéristiques : l'étendue et la capacité de penser. Le corps est étendu et ne pense pas, et l'esprit pense, mais n'est pas étendu. Descartes écrit dans les *Méditations métaphysiques* :

[...] mon essence consiste en cela seul, que je suis une chose qui pense, ou une substance dont toute l'essence ou la nature n'est que de penser. Et quoique peut-être (ou plutôt

---

<sup>19</sup> Toutes ne pourront bien entendu pas être abordées, et nous nous concentrerons sur les principales grandes familles de solutions. L'idée est de donner au lectorat une idée de la portée du problème de l'exclusion.

certainement, comme je le dirai tantôt) j'aie un corps auquel je suis très étroitement conjoint ; néanmoins, parce que d'un côté j'ai une claire et distincte idée de moi-même, en tant que je suis seulement une chose qui pense et non étendue, et que d'un autre j'ai une idée distincte du corps, en tant qu'il est seulement une chose étendue et qui ne pense point, il est certain que ce moi, c'est-à-dire mon âme, par laquelle je suis ce que je suis, est entièrement et véritablement distincte de mon corps, et qu'elle peut être ou exister sans lui. (p. 187)

Chez Descartes, le concept de Dieu est utilisé à plusieurs reprises pour justifier certaines étapes du raisonnement, et sa version du dualisme des substances est donc quelque peu désuète dans le paradigme actuel. De plus, sa position implique la possibilité d'une existence désincarnée, hors du corps, ce qui est problématique, quoique cohérent avec les incitatifs cosmogoniques en vigueur à l'époque. Toutefois, des versions plus récentes laissent de côté les justifications divines et l'immortalité de l'âme pour se concentrer sur les conditions de possibilité d'une grille de lecture dualiste dans l'état actuel des connaissances. Edward Jonathan Lowe propose dans un article publié en 2006 un dualisme non cartésien, qu'il affirme être compatible « with all that is currently known about the operations of the human central nervous system, including the brain » (Lowe 2006, p. 6). Dans son texte, Lowe met l'accent sur la distinction entre ce qu'il appelle l'acte de choix d'une part, et l'évènement neuronal d'autre part. Chacun joue un rôle distinct dans l'explication causale de l'action d'un agent.

Le dualisme est aussi une position prise au sérieux par certains physiciens, comme Henry Pierce Stapp, qui a recours au changement de paradigme survenu avec l'avènement de la physique quantique pour expliquer des notions comme le libre arbitre, le choix conscient ou encore l'action de sondage. Dans une série d'articles publiée dans les années 2000, il accuse les neurosciences et la philosophie de l'esprit de ne pas tenir compte des avancées faites par la physique contemporaine :

In spite of this seemingly relevant twentieth century development in physics, contemporary neuroscience and philosophy of mind continue to base their quests to understand consciousness on an essentially nineteenth century conceptualization of the human brain, ignoring the facts that the older conception of reality has been known to be false for almost a century, and that, in stark contrast to the nineteenth century conceptualization, contemporary orthodox physics has specified dynamical connections between brains and minds built intrinsically into it. (Stapp 2006, p. 122)

Stapp part d'une série d'observations (la physique classique est basée sur des principes erronés et la relation corps/esprit doit être étudiée à la lumière des principes de la physique quantique) et tente de montrer que l'esprit humain, qu'il distingue explicitement dans son article du cerveau et de ses états neurologiques, joue un rôle causal qui lui est propre et qu'il influence directement le monde physique duquel il est pourtant distinct. Stapp écrit : « we human beings do in fact inject our conscious intentions into the physically described world whenever we act intentionally » (Stapp 2006, p.126).

Rentrer dans les détails de la proposition de l'auteur nous éloignerait trop de notre propos, nous souhaitons simplement montrer que le dualisme est une position toujours envisagée aujourd'hui. Toutefois, les discussions contemporaines en philosophie de l'esprit se font en majeure partie au sein du paradigme physicaliste, et favorisent donc le monisme de substance. Les prochaines solutions que nous envisageons sont presque toutes des solutions monistes.

### 1.3.2 Épiphénoménisme

Une autre solution consiste à rejeter [Efficacité causale du mental] et à accepter que les propriétés mentales, bien que présentes et irréductibles à leurs corrélats physiques, n'ont pas de pouvoir causal et n'ont donc pas à être prises en considération lors de la relation causale entre P et P\*. On parle alors de propriétés *épiphénoménales* : elles existent mais ne sont pas efficaces causalement. Le défi principal d'une telle position est de justifier l'existence des propriétés en question. En effet, il est coutumier d'utiliser l'efficacité causale comme critère lorsque l'on considère la pertinence d'ajouter un élément à notre ontologie. De plus, des principes comme le rasoir d'Ockham<sup>20</sup> et la simplicité explicative qu'il suggère nous poussent à préférer une ontologie épurée quand cela est possible. Pourquoi alors accepter des entités ne jouant aucun rôle causal, ne faisant aucune différence dans le monde ? Il semble que le fardeau de la preuve soit sur les épaules de l'épiphénoméniste. Nous allons présenter brièvement la version dualiste de l'épiphénoménisme, que nous appellerons « épiphénoménisme fort », puis la version fonctionnaliste proposée par Helen Beebe, et que nous nommerons « épiphénoménisme faible ».

La forme traditionnelle de l'épiphénoménisme est une extension du dualisme, en cela que les propriétés épiphénoménales n'influencent pas le domaine physique, et n'en font donc pas partie. On en retrouve une première version au XIX<sup>e</sup> siècle, chez le philosophe Thomas Henry Huxley. À l'époque de l'auteur, le dualisme était la position la plus répandue, comme le physicalisme l'est aujourd'hui. Sa théorie prenait donc racine dans le dualisme, et la question était de savoir si et comment le mental pouvait avoir une influence sur un domaine dont il était pourtant distinct (le domaine physique). Une fois que l'on a postulé la séparation des propriétés mentales et physiques, il est en effet possible de déclarer les premières causalement inertes. Les épiphénoménistes dualistes conçoivent donc les propriétés mentales comme sans impact sur le domaine physique. Dans ce cas-ci, les problèmes que rencontrent les épiphénoménistes s'ajoutent à ceux qu'un dualiste doit régler ; il ne s'agit plus seulement de justifier l'indépendance ontologique du mental, il faut en justifier l'impotence. Dans un tel cas, avoir recours au mental pour rendre compte de

---

<sup>20</sup> Selon William d'Ockham, à succès explicatif égal, une théorie est préférable à une autre si elle est plus simple – c'est-à-dire si elle contient moins d'éléments, de principes ou encore d'étapes dans sa démonstration. Ce principe a été énormément discuté et peut être formulé de multiples façons.

l'occurrence de quoi que ce soit est complètement non pertinent, c'est pourquoi nous nommons cet épiphénoménisme « épiphénoménisme fort ». Un des intérêts d'une telle position est de tenter de concilier déterminisme et sensation de contrôle sur nos actions : notre vie mentale, les impressions que nous laissent nos expériences sensibles, ne seraient que des propriétés épiphénoménales, survenant sur ou dépendant d'une manière ou d'une autre de processus biologiques, chimiques et physiques dans nos cerveaux, sans toutefois rien causer.<sup>21</sup>

Toutefois, les épiphénoménistes ne conçoivent pas nécessairement les propriétés mentales comme complètement détachées causalement du domaine physique, et certain-e-s tentent même de formuler une version de cette thèse qui soit compatible avec le paradigme physicaliste. Il est en effet possible de concevoir les propriétés mentales comme étant causées par les propriétés physiques, et comme appartenant au même domaine que celles-ci, tout en continuant d'affirmer qu'elles ne jouent pas de rôle causal dans la production des comportements. Helen Beebe propose plutôt que les propriétés mentales jouent des rôles explicatifs au regard de l'histoire causale des événements. L'idée est de distinguer explication causale et relation causale. Une relation causale lie une cause à son effet et est exprimée par un énoncé de la forme « C cause E », où les deux éléments dont il est question sont deux événements unis par un lien causal l'un avec l'autre. Une explication causale (qui peut prendre la forme « E parce que C », pour reprendre l'exemple de Beebe) ne fait que donner de l'information pertinente sur l'histoire causale de l'évènement. Les éléments présents dans une explication causale ne sont pas nécessairement la cause ou l'effet eux-mêmes. Comme le note l'autrice : « Not all causal explanations are reports of causation: the explanans of a causal explanation need not stand to the explanandum as cause to effect » (Beebe 2004, p.301). La distinction est subtile mais permet d'apporter des réponses à un certain nombre de problèmes que pose l'étude des relations causales.

Dans l'article de 2004 que nous venons de citer, elle utilise la distinction pour clarifier la question de la causalité par omissions.<sup>22</sup> Toutefois, dans d'autres textes, elle s'intéresse à la question des propriétés mentales en ayant recours au même type d'argument. Elle affirme que malgré une critique souvent proférée à l'endroit de l'épiphénoménisme, il ne représente pas une contradiction théorique ou une absurdité aussi évidente qu'on pourrait le croire. Elle écrit dans une contribution à un ouvrage collectif qu'elle a publié en 2017 :

---

<sup>21</sup> Une telle position est cependant problématique, car il semble que les impressions que nous laissent nos expériences influencent fréquemment nos délibérations ultérieures et les comportements qui en résultent ; les présenter comme causalement inertes est donc difficile à défendre.

<sup>22</sup> Face à la présence d'omissions ou d'absences dans le discours causal de tous les jours (mon oubli d'arroser ma plante a causé sa mort), Helen Beebe accuse le sens commun de commettre une erreur en attribuant un rôle causal aux absences. Pour elle, ce dernier prendrait les explications causales pour des relations causales. Voir à ce sujet Beebe (2004).

[...] how bad would epiphenomenalism be, really? Not nearly as bad as one might think, in fact [...] the kind of epiphenomenalism that the Causal Role Problem points to, while it does undermine the causal relevance of mental properties, leaves much of what we want to say about the mental intact: the causal irrelevance of the mental does not entail that mental properties are explanatorily redundant, nor does it entail that mental events are causally inefficacious. Finally, it does not entail that conceiving the world in terms of mental properties is useless for the purposes of controlling ourselves and others. (Beebe 2017, p.301)

Ici, Beebe propose d'accepter que les *propriétés* mentales soient non pertinentes causalement, tout en conservant l'efficacité causale des *événements* mentaux. Pour elle, les propriétés mentales peuvent jouer un rôle dans les explications causales, en cela qu'on peut les utiliser pour décrire de façon précise les événements qui sont impliqués dans les relations causales (qui en sont les *relata*) ; mais ce ne sont pas elles-mêmes des causes. Elles sont juste des épiphénomènes survenant sur, ou dépendant de propriétés physiques – propriétés physiques qui elles, *sont* pertinentes causalement au sens où elles sont les propriétés responsables de la relation causale entre un événement et son effet. Un événement est qualifié de « mental » dès lors qu'il exemplifie une propriété mentale. Toutefois, celle-ci n'est pas efficace causalement. L'évènement doit aussi exemplifier une propriété physique qui fait le travail causal. Puisque cette forme d'épiphénoménisme ne conçoit pas le mental comme complètement détaché du physique, nous le nommons « épiphénoménisme faible ».

Bien qu'il existe des défenses de l'épiphénoménisme, à l'instar de celle de Beebe, il s'agit d'une position assez minoritaire en philosophie de l'esprit. Pour cette raison, nous ne l'explorerons pas plus avant. Il nous importait simplement de souligner qu'elle était défendue par certain·e·s philosophes. Tournons-nous à présent vers le fonctionnalisme, qui demeure une des positions les plus en vogue, et aux formes les plus variées.

### 1.3.3 Fonctionnalisme : un cas de non-réductionnisme ?

À la fin du XX<sup>e</sup> siècle, le fonctionnalisme s'est imposé comme la position dominante en philosophie de l'esprit. Avant son avènement, la discipline était dominée par une querelle entre béhaviorisme et théorie de l'identité. Le béhaviorisme<sup>23</sup> est une position selon laquelle l'objet d'étude de la psychologie ne devrait pas être

---

<sup>23</sup> Le terme « béhaviorisme » est utilisé pour désigner plusieurs courants de pensée rattachés à diverses disciplines. Parmi les courants les plus célèbres et les plus influents, on peut citer les travaux de John Watson et son béhaviorisme méthodologique (un de ses ouvrages majeurs est paru en 1923 et s'intitule *Behaviorism*). Plus tard au XX<sup>e</sup> siècle, Burrhus Frederic Skinner, à partir des travaux de Watson et de ceux d'Ivan Pavlov, a développé un programme de recherche et une théorie qui ont longtemps été très influents en psychologie et en psychiatrie : le béhaviorisme psychologique (l'œuvre de Skinner est très vaste mais on peut mentionner son ouvrage de 1974 intitulé *About Behaviorism*). En philosophie de l'esprit, le béhaviorisme est considéré davantage comme une attitude sceptique à l'égard de la conception des états mentaux comme des états internes d'un individu, et consiste à les considérer plutôt comme des dispositions : des tendances à agir de telle ou telle façon.



les états mentaux mais les comportements des individus. Cette tendance se développe au début du XX<sup>e</sup> siècle et prend progressivement de l'importance. Toutefois, les années 50 voient naître une théorie rivale, selon laquelle les états mentaux sont identiques à des états neuronaux.<sup>24</sup> Pendant une dizaine d'années, des partisans des deux théories débattent pour savoir si la philosophie de l'esprit devrait se concentrer sur les comportements ou sur les états neuronaux. Dans les années 60, cependant, Hilary Putnam propose de s'intéresser davantage à la *fonction* que remplit un état mental, et entreprend de répondre aux deux théories. Il écrit : « we identify organisms as in pain, or hungry, or angry, or in heat, etc., on the basis of their *behavior*. But it is a truism that similarities in the behavior of two systems are at least a reason to suspect similarities in the functional organization of the two systems, and a much *weaker* reason to suspect similarities in the actual physical details » (Putnam 1967, p. 437). Il conteste la thèse de l'identité des types qui selon lui manque sa cible en s'intéressant à la *nature* des états mentaux. Sur un autre front, bien qu'il concède aux behavioristes que l'observation du comportement est d'un certain intérêt pour l'étude de l'esprit, il n'entend pas en faire sa préoccupation principale. Pour lui, il s'agit d'une bonne raison d'adhérer à la thèse fonctionnaliste : des ressemblances de comportements entre des individus de plusieurs espèces, malgré des structures physiologiques différentes, semblent indiquer une fonction similaire, nécessaire à l'organisme, remplie chez ces individus par des processus physiologiques différents.

L'originalité de l'approche fonctionnaliste réside dans le changement de perspective qu'elle propose : on ne s'intéresse plus à la composition des états mentaux, à leur nature, mais à la fonction qu'ils remplissent. Alors que certaines tentatives pour comprendre l'esprit humain se concentrent sur la question du statut ontologique des propriétés mentales, le fonctionnalisme laisse celle-ci de côté pour s'intéresser à la question de la causalité.<sup>25</sup> Ainsi, plutôt que de se demander à quel domaine appartiennent les états mentaux, il s'agit de se concentrer sur leur rôle causal. Pour les fonctionnalistes, avoir une propriété mentale ou un état mental signifie simplement être dans un état interne qui joue un certain rôle causal. Nul besoin de préciser la nature de l'état en question. C'est le rôle causal qu'il importe de définir. Putnam écrit d'ailleurs : « I shall, in short, argue that pain is not a brain state, in the sense of a physical-chemical state of the brain (or even the whole nervous system), but another kind of state entirely. I propose the hypothesis that pain, or the state of being in pain, is a functional state of a whole organism. » (1967, p. 433). Le fonctionnalisme est souvent vu comme une des nombreuses déclinaisons du non-réductionnisme, en cela que malgré sa compatibilité avec une position physicaliste, il n'admet pas que les propriétés mentales soient réduites aux propriétés physiques. C'est même

---

<sup>24</sup> Plus précisément, la thèse en question affirme que chaque *type* d'état mental est identique à un *type* d'état neuronal. Les thèses de l'identité sont partagées entre identité des types et identités des occurrences (*token identity*). Nous présentons cette distinction à la dernière section de ce chapitre et l'examinons plus avant dans le second chapitre.

<sup>25</sup> L'ontologie et l'étude de la causalité sont parfois considérées comme deux objets d'étude de la métaphysique.

une des intuitions motivant sa genèse. Toutefois, en raison de sa neutralité sur la question ontologique de la nature des états mentaux, certain·e·s philosophes affirment que le fonctionnalisme est compatible avec des positions dualistes ou encore, nous le verrons dans le troisième chapitre, avec une version de la théorie des tropes.

Le fonctionnalisme étant une des positions les plus populaires en philosophie depuis les années 70, il en existe une grande variété et à part la définition sommaire que nous venons d'en faire, certaines d'entre elles n'ont pas grand-chose en commun. Dans son livre *Philosophy of Mind*, paru en 1996 puis réédité en 2006 et en 2010, Kim distingue deux grandes familles de positions fonctionnalistes. Cette division apparaît chez plusieurs philosophes sous des noms différents, mais nous les nommerons respectivement « théories du rôle fonctionnel » et « théories de la spécification fonctionnelle (*realizer-functionalism*) ».

La théorie de Putnam est un exemple classique de théorie du rôle fonctionnel. Dans ce type de théorie, les propriétés mentales sont des propriétés d'ordre supérieur, et ne remplissent pas directement les rôles fonctionnels qui les caractérisent. Plutôt, avoir une propriété fonctionnelle (comme une propriété mentale par exemple), consiste à être dans un certain état (cérébral, dans le cas des humains) qui joue un rôle causal spécifique. Mais les propriétés mentales, elles, sont « clearly distinct from the physical mechanisms that fill the role » (Kim 2010, p. 187). D'où l'expression « ordre supérieur » : avoir une propriété mentale, pour un système, consiste à être dans un certain état – à posséder une autre propriété d'ordre inférieur qui remplit un certain rôle causal.<sup>26</sup>

Pour les théories de la spécification fonctionnelle, en revanche, nul besoin d'avoir recours à des propriétés d'ordre supérieur : ce type de théorie fonctionnaliste doit simplement décrire des rôles fonctionnels, et trouver les propriétés qui les remplissent : les propriétés fonctionnelles sont ces propriétés (au lieu d'être de nouvelles propriétés, d'ordre supérieur). La théorie de Lewis appartient à cette catégorie (Lewis 1966).

Un des attraits du fonctionnalisme est sa capacité à expliquer comment des créatures ayant des structures physiologiques très différentes peuvent partager des états semblables. Un bon exemple de succès explicatif de la théorie est le cas de la douleur : êtres humains et chats ne partagent pas beaucoup de traits physiologiques, et leurs cerveaux respectifs n'ont pas la même taille ni la même structure, mais tous deux réagissent souvent de façon assez semblable lorsqu'exposés à un stimulus douloureux. Cette affirmation conduit au problème de la réalisabilité multiple, sur lequel nous reviendrons à plusieurs reprises dans ce

---

<sup>26</sup> Nous utilisons les expressions « ordre supérieur » et « ordre inférieur », mais Kim emploie « premier ordre » et « second ordre ». Ces termes ne renvoient pas à un ordre ontologique. Comme le note Kim : « no special meaning needs to be attached to the terms “first-order” and “second-order”; these are relative terms » (Kim 2010, p.183).

mémoire car il est un des points de discordance principaux entre les réductionnistes comme Kim et les fonctionnalistes. Pour ces derniers, postuler l'existence d'une fonction similaire qui est susceptible d'être réalisée par des processus physiologiques différents permet d'éviter d'avoir à affirmer qu'il n'y a pas d'état commun, appelé « douleur », que partageraient chats, humains et autres espèces. En cela, il s'agit d'un succès explicatif évident du fonctionnalisme. Un autre avantage de la position est sa neutralité, que l'on a déjà mentionnée, à propos de la question de la nature des états mentaux : dualistes et physicalistes peuvent donc y adhérer en conservant leurs ontologies respectives (bien que les fonctionnalistes se décrivent pour la plupart comme des physicalistes).

De plus, au sein même du physicalisme, il semble possible pour un fonctionnaliste d'adhérer à une position réductionniste ou à une position non réductionniste. Le réductionnisme affirme que l'on peut réduire les propriétés mentales à des propriétés physiques, alors que le non-réductionnisme souhaite, malgré sa volonté de s'inscrire dans un cadre physicaliste, maintenir une certaine indépendance des propriétés mentales par rapport aux propriétés physiques. Puisque le fonctionnalisme, on l'a vu, ne s'intéresse pas à la nature des états mentaux, il semble à première vue compatible avec les deux positions. Mais la seconde est-elle vraiment envisageable ? Pour Kim, le non-réductionnisme fait face à un problème de taille : sa volonté de conserver [Efficacité causale du mental] et [Irréductibilité] le pousse en fait à adopter une ontologie dualiste, et le conduit donc aux mêmes problèmes que le dualisme des substances exposé plus tôt. Selon Kim, il est donc difficile de se prétendre physicaliste tout en demeurant neutre sur la question ontologique, à moins d'abandonner [Efficacité causale du mental] (ce que ne semblent pas vouloir faire les fonctionnalistes). Il note dans un article intitulé « The Myth of Nonreductive Materialism » :

I will claim that a physicalist has only two genuine options, eliminativism and reductionism. That is, if you have already made your commitment to a version of physicalism worthy of the name, you must accept the reducibility of the psychological to the physical, or, failing that, you must consider the psychological as falling outside your physicalistically respectable ontology. Of course, you might decide to reconsider your commitment to physicalism; but I will not here consider what nonphysicalist alternatives there might be which are still live possibilities for us. So if I am right, the choices we face concerning the mind-body problem are rather stark: there are three — antiphysicalist dualism, reductionism, and eliminativism. (Kim 1993b, p. 267)

Si Kim a raison, et si nous ne voulons pas aller vers un dualisme des substances, il semble donc qu'il faille s'en remettre au réductionnisme ou à l'éliminativisme.<sup>27</sup> Avant de présenter le réductionnisme, qui va

---

<sup>27</sup> Une solution pour sauver le fonctionnalisme de l'écueil dualiste serait de l'arrimer à une ontologie résolument moniste. La philosophe Ann Whittle met de l'avant cette solution et associe théorie des tropes et fonctionnalisme. C'est sur cette possibilité que nous revenons dans le troisième chapitre.

nous intéresser pour le reste de ce mémoire, arrêtons-nous sur cette autre possibilité, une dernière solution potentielle au problème de l'exclusion causale.

### 1.3.4 Matérialisme éliminativiste

Le matérialisme éliminativiste est une position défendue principalement par Patricia et Paul Churchland.<sup>28</sup> Elle consiste à remettre en question l'existence des états mentaux et la pertinence d'avoir un lexique théorique complexe pour y référer. Les Churchland avancent que la psychologie du sens commun – soit la conception populaire des phénomènes mentaux, de l'action et du fonctionnement de l'esprit humain – est une théorie, et qu'elle est fautive. Selon les éliminativistes, lorsque les neurosciences seront suffisamment développées, elles relègueront cette conception classique des états mentaux au rang de théorie fallacieuse, comme l'aurait fait la chimie moléculaire pour l'alchimie si cette dernière avait subsisté assez longtemps pour y être confrontée.

Une des prémisses de la position éliminativiste est l'idée selon laquelle la psychologie populaire est une théorie.<sup>29</sup> Elle possède son propre champ lexical, son propre vocabulaire spécialisé. Comme toute théorie, elle a ses propres termes techniques : « attitude propositionnelle », « intentionnalité », etc.<sup>30</sup>. Paul Churchland précise par ailleurs que la signification de ces termes techniques *dépend* du réseau conceptuel de la théorie qu'est la psychologie populaire. La psychologie a une fonction précise : celle d'expliquer et de prédire le comportement des individus. De plus, elle tente de résoudre des énigmes posées par l'esprit humain comme la notion d'introspection, de mémoire, ou encore le problème de l'esprit d'autrui. Pour Churchland, cela est aussi un critère lui permettant d'affirmer que la psychologie populaire est une théorie.

Les éliminativistes pensent que cette théorie est fautive, et qu'elle va un jour être remplacée par des neurosciences parvenues à maturité, qui se seront partagé les tâches d'investigation et de description du fonctionnement et de la nature de l'esprit humain. Deux méthodes se retrouvent fréquemment dans les écrits des Churchland pour appuyer cette conjecture. Loin d'être convaincu·e·s par les soi-disant réussites de la psychologie populaire, qui sont à leurs yeux une des raisons principales de son inertie conceptuelle, Patricia et Paul proposent de se pencher aussi sur ses échecs. Cela leur fournit leur premier argument : un certain nombre de problèmes posés par l'esprit humain ne trouvent pas de solution ou même d'hypothèses d'explication satisfaisantes dans la psychologie populaire (les troubles mentaux, le sommeil, les processus

---

<sup>28</sup> Les deux philosophes ont beaucoup publié sur le sujet au fil des ans, nous nous sommes notamment inspirés de deux articles pour écrire cette section : l'un écrit par Paul en 2005, et l'autre par Patricia en 1988.

<sup>29</sup> Cette affirmation ne fait pas l'unanimité : pour les défenseurs de la théorie de la simulation, par exemple, la psychologie du sens commun se rapporte davantage à du cas par cas qu'à une théorie à proprement parler.

<sup>30</sup> L'intentionnalité est le fait de porter sur quelque chose, et est une propriété spécifique à un certain type d'états mentaux : les attitudes propositionnelles.

d'apprentissage). En se penchant sur l'histoire de la psychologie populaire, les éliminativistes font deux autres constats. D'abord, il semble que son champ d'expertise a grandement diminué : autrefois, on prêtait désirs et croyances aux vents, aux volcans et même au soleil, et on prétendait pouvoir en expliquer le comportement grâce à des attitudes propositionnelles. Ensuite, même dans son champ d'expertise actuel, la psychologie populaire ne semble pas être une théorie très dynamique. Les Churchland jugent en effet que peu de progrès ont été faits depuis l'Antiquité, et reprochent à la psychologie populaire de stagner. Enfin, le manque d'intégration de celle-ci avec des théories de domaines connexes leur permet d'affirmer que contrairement à d'autres disciplines qui font des efforts pour répondre aux exigences interdisciplinaires qui semblent cruciales à la progression du savoir, la psychologie populaire n'est pas cohérente avec les avancées d'autres disciplines.

Ces quelques arguments sont parmi les seuls disponibles pour les éliminativistes, et leur position est souvent considérée comme radicale et peu pertinente par le reste de la communauté se penchant sur la question du statut des états mentaux. Les objections à la thèse éliminativiste sont nombreuses, et une des principales est son manque de pertinence dans l'état actuel des connaissances en neurosciences. En effet, même si les éliminativistes ont raison et que des neurosciences matures remplacent un jour la psychologie, la psychiatrie, et jusqu'au jargon populaire utilisé pour parler de l'esprit, cela n'est pas pour tout de suite, et quand bien même des découvertes allant en ce sens commenceraient à être faites dans les prochaines décennies, il y a fort à parier que le remplacement d'un pan entier du langage populaire ne se ferait pas d'un seul coup.

Une autre critique de l'éliminativisme souvent mise de l'avant par ses détracteurs est qu'elle ne prend pas au sérieux les succès théoriques et pratiques de la psychologie populaire. Notre capacité à prévoir avec précision les réactions ou les motivations d'individus, au moins lorsqu'il s'agit d'actions simples, semble être une bonne raison de croire que nous ne nous trompons pas lorsque nous nous attribuons mutuellement des désirs et des croyances. La réponse des éliminativistes est alors souvent de prendre en exemple d'autres théories erronées ayant joui très longtemps d'un succès explicatif (comme l'*Almageste* de Ptolémée) pour modérer l'efficacité de cette critique.

Tant en raison de ces arguments que de ces critiques, le bien-fondé de la position éliminativiste peut sembler difficile à saisir. Toutefois, tout comme les autres solutions potentielles au problème de l'exclusion, elle a le mérite d'alimenter la réflexion au sujet du mental et de son rapport au physique. Ainsi, le recensement effectué dans les pages précédentes a montré, nous l'espérons, la pertinence et la plus grande force du problème de l'exclusion : il permet d'établir un cadre de réflexion fructueux pour des questions métaphysiques et épistémologiques à propos de l'esprit ; des questions concernant la nature des propriétés mentales, leur

rôle dans la production des comportements, etc. Avant de passer au chapitre suivant et de nous intéresser plus en détail à la théorie des tropes, nous allons présenter une dernière famille de solutions à laquelle celle-ci appartient.

### 1.3.5 Réductionnisme

On qualifie de « réductionniste » une thèse qui propose de réduire des entités, des termes techniques, voire des théories complètes à d'autres. Bien qu'il existe des positions réductionnistes à propos de plusieurs domaines, nous nous concentrerons ici sur la réduction du mental au physique. Nous allons répertorier différentes formes de réductionnisme. Il va d'abord s'agir de distinguer réductionnisme ontologique et réductionnisme épistémologique, deux familles de thèses réductionnistes dont les membres portent sur plusieurs domaines différents. Puis, sur la question des états mentaux, nous distinguerons deux composantes du réductionnisme : l'identité des types (*type-type identity*) et l'identité des occurrences (*token identity*). Nous parlerons aussi de l'objection majeure au réductionnisme à propos du mental (l'argument de la réalisabilité multiple), objection qui, selon Kim, explique l'immense popularité du non-réductionnisme en philosophie de l'esprit (Kim 1992, p. 310).

Commençons par distinguer deux types d'affirmations ou de thèses réductionnistes. Kim donne la définition suivante dans l'article « The Nonreductivist's Trouble with Mental Causation »<sup>31</sup>: « the distinctive properties of entities at a given level are *reducible to, or reductively explainable in terms of,* the properties and relations characterizing entities at lower levels. That is reductionism » (Kim 1993b, p. 338). Il semble y avoir deux notions ici. Si les propriétés mentales, par exemple, sont elles-mêmes réduites aux propriétés physiques, alors la thèse affirme quelque chose à propos de ce qui existe : à savoir que les entités que l'on désigne à l'aide du champ lexical du mental sont en fait identiques à des entités physiques. C'est un réductionnisme *ontologique*. C'est cette thèse qui constitue la position de Kim, à plusieurs nuances près.<sup>32</sup> Cette position est très contestée par les non-réductionnistes, qui ne souhaitent pas abandonner l'indépendance ontologique du mental par rapport au physique, et affirment que les réductionnistes risquent de sombrer dans l'éliminativisme.

L'autre type d'affirmation réductionniste, un peu moins forte que la réduction ontologique, apparaît également dans la définition de Kim. En effet, la deuxième expression employée par l'auteur dans l'extrait cité au paragraphe précédent ne fait référence qu'à l'*explication* donnée d'une succession d'évènements. Dans ce cas-ci, le réductionnisme (à propos du mental, par exemple) consiste à affirmer qu'il est possible d'expliquer

---

<sup>31</sup> Cet article a été publié par Kim en 1993, dans un recueil de ses anciens travaux, mais est paru la même année dans l'ouvrage collectif *Mental Causation*, dirigé par John Heil et Alfred Mele.

<sup>32</sup> Dans *Physicalism, or Something Near Enough*, Kim concède par exemple qu'il ne croit pas possible de réduire les états phénoménaux – les *qualia* – de la même façon que le reste des états mentaux.

des événements impliquant des propriétés habituellement associées au mental à l'aide de termes physiques, et ce sans perte de contenu cognitif entre les explications. On parle de réduction d'une explication (ou d'une théorie explicative) à une autre en vertu du fait qu'il serait possible de donner deux (ou plusieurs) explications du même phénomène, faisant intervenir des termes théoriques et des lois ou règles différentes, sans perdre de contenu ni d'information causale pertinente au sujet de ce phénomène lorsque l'on passe de la première explication à celle à laquelle elle est réduite. Appelons cette version plus faible<sup>33</sup> le réductionnisme épistémologique. Pour en saisir l'idée générale, on peut partir d'une question formulée par Kim : « the question whether the same bit of behavior, say an action we perform such as climbing a ladder, can be given both a "mechanistic" explanation, in terms of physiological processes and laws, and a "purposive" explanation, in terms of "reasons" (e.g., goals and beliefs) » (Kim 1989a, p. 237). La réponse à cette question peut servir à distinguer réductionnistes et éliminativistes, mais aussi réductionnistes et non-réductionnistes. Selon la version classique du réductionnisme épistémologique (Nagel, 1961), les explications en termes mentaux peuvent être ramenées à des explications en termes physiques grâce à des lois-ponts (*bridge laws*) qui unissent les prédicats de la théorie psychologique à des prédicats (éventuellement complexes) de la théorie neurophysiologique à laquelle elle est réduite. En pareil cas, la théorie réduite et ses concepts sont conservés. Le réductionnisme épistémologique n'est donc pas une forme d'éliminativisme. Cependant, les explications psychologiques ne sont pas considérées comme indépendantes des explications physiques, et ce sont ces dernières qui donnent le portrait le plus détaillé et théoriquement le plus fécond des phénomènes mentaux. Les non-réductionnistes, on s'en doute, s'opposent à l'idée qu'il est possible de formuler de telles lois de connexion entre le lexique des états mentaux à celui des états neurophysiologiques.

Il est important de mentionner que les déclinaisons en philosophie de l'esprit de ce qu'on a appelé « réductionnisme ontologique » sont fréquemment appelées « théories de l'identité » dans la littérature. La distinction que nous proposons entre réductionnisme ontologique et épistémologique pourrait donc être conçue comme une distinction entre les théories de l'identité (des thèses ontologiques selon lesquelles les propriétés mentales sont identiques à des propriétés physiques) et les théories de l'explication (des thèses épistémologiques selon lesquelles il est possible de donner une explication en termes physiques d'événements impliquant des états mentaux). Toutefois, la formulation disjonctive faite par Kim et citée plus haut (*reducible to, or reductively explainable in terms of*), nous motive à concevoir ces deux familles de thèses comme des réductionnismes. Il est par ailleurs intéressant de se demander si les deux vont forcément ensemble.

---

<sup>33</sup> Elle est plus faible car elle n'implique pas de réduction ontologique, en tout cas pas au premier regard.

Il semble assez évident qu'un réductionnisme ontologique (la réduction d'une entité à une autre) implique un réductionnisme épistémologique (la réduction d'une théorie décrivant les comportements d'une entité à une autre). Imaginons deux théories A et B portant chacune sur un ensemble d'entités. S'il s'avère que les éléments décrits par A sont en fait tous identiques à certains éléments décrits par B, ou constitués d'éléments décrits par B, alors il semble cohérent de penser que la théorie A peut être réduite à la théorie B – c'est-à-dire qu'il est possible d'expliquer tout évènement impliquant des éléments décrits par la théorie A en ne faisant appel qu'à B. Le fait que l'inverse soit aussi vrai (qu'un réductionnisme épistémologique entre deux théories implique un réductionnisme ontologique des entités qu'elles décrivent) semble être suggéré par Kim dans *Philosophy of Mind*.<sup>34</sup> L'implication réciproque de ces deux thèses présente un avantage important : si elles sont équivalentes, il suffit de démontrer que l'une d'entre elles est vraie pour s'assurer que la seconde le soit également. Si l'on veut défendre le réductionnisme ontologique d'un domaine à un autre, mais qu'on est incapable de le démontrer à l'aide de recherches empiriques, une solution argumentative peut être de démontrer l'équivalence des réductionnismes ontologique et épistémologique et de démontrer que le second s'applique entre les domaines qui nous occupent. Toutefois, et c'est l'une des conclusions de Kim, la *possibilité* d'un réductionnisme épistémologique ne devrait pas être un incitatif à abandonner – ou à *éliminer* – la théorie réductible.<sup>35</sup> C'est là le point de désaccord crucial entre réductionnistes et éliminativistes. Cette précision est importante afin de bien circonscrire notre propos, et nous croyons que ces remarques de Kim sont capitales. Même si l'on est capable d'établir une réduction ontologique du mental au physique, cela ne devrait pas être une motivation à abandonner ou à rejeter avec dédain la psychologie, son indépendance et sa pertinence.

La théorie des tropes, nous le verrons, offre comme solution au problème de l'exclusion un réductionnisme ontologique, et implique donc la possibilité d'un réductionnisme épistémologique. En effet, elle fait des affirmations à propos de la réalité et de l'existence de certaines entités, les tropes.

---

<sup>34</sup> Dans la section intitulée « Closing the Explanatory Gap : Reduction and Reductive Explanation », il affirme que lorsqu'on s'intéresse au problème de la conscience (aussi appelé « *hard problem* », ou « problème du fossé explicatif » dans la littérature) en termes de réductionnisme explicatif, on le fait car on conçoit la possibilité d'une réduction épistémologique sans réduction ontologique : « someone who favors this approach to the explanatory gap is apt to do so because of the thought that reductive explanation is possible even where reduction is not » (Kim 2010, p.311). Kim discute de cette position dans la section et semble plutôt en prendre le contrepied.

<sup>35</sup> Au sujet de la psychologie, Kim mentionne qu'il considère que le moyen de la sauver serait d'arrêter de la considérer comme une *théorie*, justement au sens où les éliminativistes l'entendent. Il écrit dans une note de bas de page : « We will do better to focus on its normative role in the evaluation of actions and the formation of intentions and decisions. If vernacular psychology competes against cognitive science in the prediction game, it cannot win, and the best thing it can hope for is reductive absorption into its more systematic (and better funded) rival » (Kim 1989a, p. 263).



Mais avant de l'aborder, penchons-nous sur une dernière distinction entre deux versions du réductionnisme ontologique à propos du mental : l'identité des occurrences<sup>36</sup> et l'identité des types. L'identité des occurrences est la thèse selon laquelle il est possible de réduire *chaque* état mental à un état cérébral. Ceci laisse ouverte la possibilité que deux états mentaux du même genre soient réduits à des états cérébraux d'un genre différent. L'identité des types, elle, affirme que chaque catégorie d'état mental est réductible à une catégorie d'état cérébral. Il est important de noter que l'identité des types implique l'identité des occurrences, mais que l'inverse n'est pas vrai. L'identité des occurrences est donc une thèse plus faible que l'identité des types (Fodor, 1974). Une des objections principales au réductionnisme, que nous devons aborder plusieurs fois dans les pages qui suivent, se base sur l'idée selon laquelle le réductionnisme est l'identité des types. Cette objection a été introduite par Putnam en 1967 dans l'article « Psychological Predicates » ; il s'agit de la réalisabilité multiple, à laquelle nous avons déjà fait allusion. Pour Putnam, le fait qu'on puisse observer des états mentaux similaires chez des individus d'espèces différentes montre qu'il existe différentes façons de réaliser le même type d'état mental. Ainsi, il ne serait pas possible d'identifier le type « douleur » avec le type « stimulation des fibres C » (pour reprendre un exemple classique de la littérature sur le sujet), car d'autres systèmes (biologiques ou non) pourraient eux aussi instancier des états mentaux s'apparentant à de la douleur sans avoir de stimulations des fibres C. En réponse à cette objection, les réductionnistes mettent de l'avant le fait que cette objection ne pose pas de problème pour l'identité des occurrences, voire pour l'identité des types à l'intérieur d'une même espèce, et que leur position n'est donc pas menacée. Bien entendu, des versions de l'argument de la réalisabilité multiple adaptées à ces thèses moins fortes de l'identité furent développées par la suite. Elles comptent aujourd'hui parmi les défis les plus sérieux auxquels fait face le réductionnisme. C'est précisément la question de l'efficacité de l'argument de la réalisabilité multiple face à la position réductionniste qui va nous intéresser dans le dernier chapitre de ce mémoire. Mais avant, il nous faut présenter en détail la version du réductionnisme ontologique qu'offre la théorie des tropes.

## 1.4 Conclusion

Dans ce chapitre, nous nous sommes servis de l'argument de l'exclusion comme cadre théorique pour brosser le portrait général d'un problème qui émerge lorsqu'on tente de concilier plusieurs de nos intuitions à propos de nos états mentaux. Nous avons exploré certaines des solutions à ce problème, mais nous avons vu qu'elles soulevaient chacune leurs propres enjeux théoriques. La solution qui nous intéresse consiste à réduire les propriétés mentales à des propriétés physiques, sans pour autant déclarer la mort de la psychologie ou d'autres discours ayant recours à des termes du champ lexical du mental, et ce afin d'éviter du

---

<sup>36</sup> « Token » est parfois traduit par « jeton ». Le terme est d'ailleurs pertinent eu égard à sa proximité sémantique avec la notion de trope. Nous revenons là-dessus dans le second chapitre.

même coup l'écueil dualiste et les dangers de l'éliminativisme : c'est le réductionnisme. Cependant, cette position est elle-même menacée par l'argument de la réalisabilité multiple. Il sera question dans les deux chapitres suivants de caractériser la version particulière du réductionnisme qui nous intéresse, puis de voir à quel point elle est efficace pour répondre aux objections auxquelles celui-ci fait face.

## Chapitre 2 : La théorie des tropes

### 2.0 Introduction

Il a beaucoup été question de propriétés dans le premier chapitre, et nous avons notamment vu que selon la conception de la causalité que nous avons retenue, les événements en causent d'autres en vertu du fait qu'ils sont des instanciations de propriétés par des objets à un moment du temps. Dans ce chapitre, l'analyse de cette notion va être au cœur de nos discussions. Cependant, nous allons laisser quelque peu de côté la question du statut des états mentaux pour nous concentrer sur les propriétés de façon plus générale, en tant qu'éléments d'une *ontologie*. Les questions ontologiques, en philosophie, sont les questions relatives à ce qui existe. Il s'agit d'une des sous-disciplines de la métaphysique. Nous allons nous pencher sur la théorie des tropes, une position ontologique visant à désambiguïser la notion de propriété et à en faire l'élément de base de la réalité. Dans un premier temps, nous allons brosser un portrait général de la notion de propriété en ontologie, définir ce que sont les tropes, et indiquer ce qui distingue la théorie des tropes d'autres positions ontologiques (2.1). Puis, nous ferons certaines précisions sur les notions de niveau de réalité et d'explication (2.2), dont nous avons parlé rapidement au premier chapitre. Finalement, nous nous intéresserons à la notion de classes de ressemblance, et au lien entre théorie des tropes et fonctionnalisme (2.3).

### 2.1 La notion de propriété : une épine ontologique

Le problème de l'exclusion, que nous avons présenté au premier chapitre, fait intervenir la notion de propriété, et la plupart des solutions à celui-ci discutent du rôle joué par les propriétés. Elles tentent de résoudre l'incohérence des thèses impliquées dans l'argument en précisant le rôle joué par certains types de propriétés. Certaines de ces solutions – les théories de l'identité – ont comme visée d'identifier des propriétés d'un certain domaine avec des propriétés d'un autre domaine. Il est donc important de bien saisir la notion de propriété et les enjeux qu'elle soulève pour être conscient des coûts et bénéfices théoriques d'une telle approche. Dans cette section, nous allons nous pencher sur certains problèmes que pose cette notion en métaphysique.

#### 2.1.1 Les propriétés : universaux ou particuliers ?

La notion de propriété est abondamment discutée en philosophie et une place de choix lui est généralement faite dans les ontologies. La question est le plus souvent de savoir quelle place elle doit occuper au juste (certaines positions prônent l'abandon de cette notion, mais nous ne nous y arrêterons pas). La question des propriétés est parfois abordée sous l'angle de la distinction entre universaux et particuliers. Les propriétés sont-elles des universaux ou des particuliers ? Existe-t-il des universaux au même sens qu'il existe des particuliers ? Aborder en détail ce débat métaphysique nous écarterait trop de notre propos, mais il est important d'en dire quelques mots pour comprendre l'implication que pourrait avoir sur lui une théorie

ontologique qui admet l'existence de tropes. Deux conceptions font donc habituellement surface lorsqu'on parle de propriété : s'agit-il, d'une part, d'un élément commun possédé par plusieurs objets, et servant à expliquer leurs ressemblances ? En ce sens, on pourrait par exemple parler de la propriété d'être rouge, possédée à la fois par mon chandail et par le cache-pot de ma plante. Ou s'agit-il, d'autre part, d'un élément particulier, unique, possédé par un seul objet ? Dans ce cas-ci, on parlerait de deux propriétés différentes : le rouge du cache-pot et le rouge de mon chandail. Dans les deux cas, il semble que la propriété doive être possédée ou instanciée par un objet<sup>37</sup>. Mais la question de fond demeure : faut-il concevoir les propriétés comme des universaux communs à plusieurs objets ou des particuliers propres à un seul ?

Pour les philosophes adeptes de la théorie des tropes, c'est précisément cette ambiguïté de la notion de propriété qui pose problème, et la régler permet de résoudre plusieurs enjeux théoriques dans différents champs de la philosophie. Dans certains textes, le terme « propriété » est utilisé pour désigner le premier élément décrit dans le paragraphe précédent. Dans d'autres, il est employé pour désigner le second. Toutefois, il est important de distinguer ces deux usages. Le constat de David Robb à ce sujet est particulièrement clair :

There are two ways to read "property". One is as "universal" or, for the nominalist, "class". On this reading, the property F is the universal F-ness or the class of all Fs. Properties in this sense are unifying entities. They are what all Fs have in common: either the Fs all share a universal or they all belong to the same class [...] But the other reading of "property" is as "abstract particular" or, as I would prefer, "trope". On this reading, properties are particulars, wholly present in the individuals that instantiate them but logically incapable of being (at the same time) wholly present elsewhere. (Robb 1997, p. 186)

Avant de revenir sur la première façon de lire le terme « propriété », et de clarifier les notions d'universel et de classe, penchons-nous sur la définition de la notion de trope donnée par Robb : les tropes sont des particuliers abstraits. L'utilisation du terme « trope » en philosophie analytique remonte à un article de Donald C. Williams paru en 1953 et intitulé « On The Elements of Being ».<sup>38</sup> Il y défend une position ontologique selon laquelle les éléments constituant la réalité sont des propriétés, qu'il conçoit comme des particuliers non des types. Après avoir caractérisé ce qu'il appelle des « composants abstraits » au début de son texte, il affirme que ceux-ci « are the primary constituents of this or any possible world, the very alphabet of being. They not only are actual but are the only actualities, in just this sense, that whereas entities of all other categories are literally composed of them, they are not in general composed of any other sort of entity »

---

<sup>37</sup> Tous les théoriciens des tropes ne sont pas d'accord avec cette affirmation. Nous y reviendrons à la section 2.1.3.

<sup>38</sup> Williams lui-même attribue l'usage du terme « trope » dans un texte philosophique à George Santayana (un philosophe du début du XX<sup>e</sup> siècle n'appartenant pas à la tradition analytique), mais précise que le sens dans lequel ce dernier l'emploie diffère quelque peu de celui qu'il souhaite mettre de l'avant, d'où la tendance à lui attribuer l'origine du terme en philosophie analytique.

(Williams 1953, p. 7). Pour lui, les tropes constituent les objets concrets qui peuplent le réel. C'est ce sens que reprennent des philosophes comme Robb.<sup>39</sup>

Pour saisir cette définition, il faut préciser ce que l'on entend par « particulier » et ce que l'on entend par « abstrait ». Nous l'avons mentionné au début de la section, la question des universaux est une question récurrente en philosophie. La notion de particulier utilisée dans la définition qui nous intéresse ici est la contrepartie de celle d'universel. Très grossièrement, un particulier est quelque chose d'unique, qui n'est pas présent à plusieurs endroits en même temps ou encore possédé par plusieurs objets en même temps, par opposition avec les universaux, qui sont justement des éléments présents dans différents objets, et qui en expliquent la ressemblance. Les tropes sont donc des particuliers, des éléments uniques qui constituent la réalité. C'est là l'idée principale de toute théorie des tropes : la réalité est constituée d'éléments particuliers, tous distincts les uns des autres.

Alors que signifie « abstrait » ? Tout comme la notion de particulier s'oppose à celle d'universel, on retrouve en métaphysique des discussions sur une distinction entre les objets concrets et les objets abstraits. Pour Williams, abstrait et universel sont à tort pris comme allant de pair, alors que l'abstraction ne signifie pas nécessairement le fait de ne pas être singulier, unique, mais plutôt d'être « partial, incomplete, or fragmentary » (Williams 1953a, p. 15). En fait, pour Williams, les tropes constituent la réalité, mais pas de la même façon que les pattes, le siège et le dossier constituent une chaise. Selon lui, les tropes sont : « the "thin" or "fine" or "diffuse" sort of constituent, like the color or shape of our lollipop, in contrast with the "thick", "gross", or chunky sort of constituent, like the stick in it » (Williams 1953a, p. 15). Puisqu'il dissocie les dichotomies abstrait/concret et universel/particulier, il est donc possible de reconnaître que les pattes de la chaise, ou le bâton de la confiserie sont des particuliers concrets alors que les tropes qui les constituent sont des particuliers abstraits, précisément en cela que l'on ne peut les saisir avec précision. Dans son article, l'auteur aborde longuement la notion d'abstraction et les critiques qui lui sont faites. Nous ne nous arrêterons pas davantage dessus, il nous faut cependant garder en tête que les tropes sont des particuliers abstraits, et qu'ils ne sont ainsi ni des universels ni des objets concrets.

Les tropes sont les éléments constitutifs de la réalité, qui constituent ce que Williams appelle « the alphabet of being » (Williams 1953a, p. 7). Dans la seconde partie de son article séminal, Williams dresse une longue liste d'exemples de tropes. Parmi elles : la texture, la forme ou encore la couleur d'une statue, une performance musicale, une chanson ou une symphonie (Williams 1953b, p. 171). L'auteur donne aussi des

---

<sup>39</sup> Williams est partisan de la théorie du faisceau ; nous en parlerons un peu plus loin. John Heil, l'autre auteur dont nous discuterons la position, a une perspective un peu différente sur laquelle nous ferons également les précisions qui s'imposent.

exemples de choses qui *ne sont pas* des tropes : la statue elle-même et les choses concrètes en général (p. 172). Toutefois, il est bon de le rappeler, il précise que les tropes *sont* sans aucun doute des parties constituantes de ces objets concrets. Parmi les autres exemples que l'on trouve dans ses articles, il mentionne que Socrate est un particulier concret, alors que sa sagesse est un particulier abstrait – un trope.

Tournons-nous à présent vers l'autre sens du terme « propriété », le premier que Robb mentionne. Selon lui, la notion est parfois comprise comme quelque chose unissant différents éléments : une propriété partagée par plusieurs objets. Par exemple, lorsque l'on parle du rouge de deux panneaux arrêts différents, on semble parler d'une propriété commune, possédée par les deux panneaux. Alors pourquoi Robb parle-t-il d'universaux *et* de classes ? Quelle est plus particulièrement la position des nominalistes à laquelle il fait allusion ? Nous l'avons mentionné rapidement un peu plus haut, les universaux sont des éléments communs à plusieurs objets. Les philosophes qui conçoivent les propriétés comme des universaux et qui considèrent que ces entités existent sont les réalistes. Les nominalistes, eux, rejettent l'existence d'entités universelles et ne postulent que des particuliers dans leur ontologie. Toutefois, ils doivent expliquer les ressemblances entre les particuliers sans pouvoir affirmer qu'une catégorie universelle existe et regroupe ces différents particuliers. Certaines théories nominalistes ont alors recours à la notion de classe de ressemblance. Il s'agit de catégories dressées *a posteriori*, à la suite d'observations empiriques ou de conceptualisations par des agents. Elles sont donc imparfaites et conditionnées par les croyances, les perceptions, le langage, et l'histoire du progrès des connaissances de ces agents, au même titre que nos meilleures théories scientifique<sup>40</sup>. Pour les nominalistes, la notion de classe de ressemblance sert donc à rendre compte de la ressemblance entre deux objets distincts, sans faire appel à des catégories universelles. Un théoricien des tropes qui ne souhaite pas introduire d'universaux dans son ontologie peut se servir de la notion de classe de ressemblance pour rendre compte de ce que certains de ces particuliers partagent.

Il n'est pas nécessaire, pour défendre une théorie des tropes, de prendre position dans le débat entre nominalistes et réalistes. Robb, par exemple, affirme que son *monisme de trope* est neutre sur la question de l'existence d'universaux, et préfère donc parler de *type* lorsqu'il parle de propriété au sens de quelque chose de commun partagé entre plusieurs objets. Il appelle sa position ainsi car il ne postule que des tropes dans son ontologie. Sa théorie est donc résolument moniste. Même si Robb considère qu'il est préférable de s'abstenir de trancher et qu'il vaut mieux parler de « types » (Robb 1997, p. 186) là où les nominalistes parlent de « classes » et où les réalistes parlent d'« universaux », nous parlerons de classes de ressemblance pour la

---

<sup>40</sup> Tel que mentionné brièvement dans le premier chapitre, nous pensons que le lien entre antiréalisme scientifique et théorie des tropes est fort, et que la pertinence de ces deux perspectives se font écho. Nous revenons plus en détail sur les raisons qui nous poussent à faire cette affirmation dans le dernier chapitre.

suite de ce mémoire lorsque nous aborderons ce que Robb appelait la première lecture de la notion de propriété.<sup>41</sup> En effet, nous allons discuter plus loin la distinction entre identité de type et identité d'occurrence. Nous allons donc réserver le terme « type » à cet usage pour ne pas rendre la lecture confuse.

La notion de propriété au sens particulier du terme doit donc être comprise, pour Robb et pour les autres philosophes dont nous discuterons les thèses dans ce mémoire, comme la notion de trope. Robb affirme dans son article qu'il se sert de la notion de trope pour répondre à un enjeu lié à l'efficacité causale du mental : le problème de l'exclusion. Avant de clore cette section sur la distinction entre les sens particulier et universel du terme « propriété », il est donc important de noter que pour Robb, et pour les philosophes qui défendent une théorie des tropes en général, les tropes sont les propriétés responsables des relations causales : « although tropes are not types (i.e. universals or classes), they fill a role traditionally assigned to types : they characterize particulars such as objects and events. And they also fill another role that types are often thought to play ; they are the “properties” that are causally relevant in causal relations ; they are, as I shall say for short, the properties of causation » (Robb 1997, p. 187). On peut déjà apercevoir la stratégie qui sera celle de Robb pour répondre au problème de l'exclusion : clarifier l'usage du terme « propriété » dans les différentes thèses qui constituent l'argument de l'exclusion. Nous revenons en détail sur cette stratégie, et sur la façon dont les tropes peuvent servir à résoudre le problème de l'exclusion en fin de chapitre. Pour l'instant, il nous faut continuer à préciser la notion de trope. Voyons à présent où situer celle-ci par rapport à une autre distinction souvent discutée en métaphysique.

### **2.1.2 Les propriétés : dispositionnelles, catégoriques ou qualitatives ?**

La notion de propriété, en philosophie, est souvent en lien avec celle de causalité : nous l'avons vu dans le premier chapitre où plusieurs des thèses du problème de l'exclusion nous poussait à aborder les deux notions conjointement. Pour Robb, nous venons de le voir, et il en est de même pour John Heil, l'autre auteur sur lequel nous allons surtout nous attarder dans ce chapitre<sup>42</sup>, les tropes *sont* les propriétés responsables des relations causales. Toutefois, une autre distinction existe au sujet de la notion de propriété : certains philosophes estiment en effet qu'il faut scinder celles-ci en deux groupes : les propriétés dispositionnelles et les propriétés catégoriques. Que signifient ces deux termes ? Où situer les tropes dans cette dichotomie ?

---

<sup>41</sup> Nous sommes toutefois d'accord avec Robb sur la question de la possibilité d'être neutre sur la question des universaux en défendant une théorie des tropes.

<sup>42</sup> Nous avons choisi de nous concentrer sur ces deux auteurs car ils s'intéressent aux implications de la théorie des tropes en philosophie de l'esprit, et leurs articles traitent notamment du problème de l'exclusion. D'où la pertinence de leurs travaux dans le cadre de ce mémoire. Toutefois, bien que nous nous concentrerons sur leur position, nous nous arrêterons aussi sur celle d'Ann Whittle dans le troisième chapitre.

Heil et Robb s'intéressent à cette distinction entre propriétés dispositionnelles et propriétés catégoriques dans leur article de 2003 : « Dispositional properties are those that characterize how an object would behave in such-and-such circumstances ; [...] Categorical properties, in contrast, are wholly here-and-now properties » (Heil et Robb 2003, p. 178). Une propriété dispositionnelle est une propriété qu'un objet manifeste si certaines circonstances sont réalisées. Parmi les propriétés dispositionnelles, on retrouve la solubilité, la malléabilité, la fragilité, etc. Ainsi, un objet manifeste qu'il est soluble en se dissolvant lorsqu'il est plongé dans un liquide. Une propriété catégorique est une propriété qu'un objet instancie du seul fait qu'il est d'une certaine façon. Parmi les propriétés catégoriques, on a la masse, la taille, la forme, etc. Les propriétés dispositionnelles sont aussi appelées des dispositions, ou des *pouvoirs*. Le lien avec l'idée de pouvoir causal est d'ailleurs assez évident : on peut dire qu'une propriété qui caractérise le comportement de l'objet qui l'instancie (ou qu'elle constitue) lui confère un pouvoir causal. Précisément celui de se comporter de telle ou telle manière dans une situation précise.

Heil et Robb s'intéressent à la distinction car ils considèrent qu'elle est largement répandue et acceptée par les philosophes, et que ceux-ci postulent souvent de surcroît que *toute* propriété est soit dispositionnelle soit catégorique, mais jamais les deux. Cette thèse est notée (D) dans leur article : « many philosophers apparently regard (D) as so obviously true that it needs no defense » (Heil et Robb 2003, p. 184). Les deux auteurs considèrent toutefois que la distinction est superflue, et qu'il faut rejeter la thèse en question. En effet, pour eux, toute propriété est à *la fois* dispositionnelle et catégorique (Heil et Robb 2003, p. 184). Une propriété ne peut pas être l'une sans être l'autre. Pour en arriver à cette affirmation, les auteurs procèdent en trois temps. Ils affirment d'abord qu'on ne peut pas concevoir de propriété purement catégorique. Si l'on veut conserver (D), il faut alors affirmer que toute propriété est exclusivement dispositionnelle. Or, ils affirment que cette position est également intenable : il faut donc rejeter (D), et affirmer que toute propriété est à la fois dispositionnelle et catégorique. Regardons ce raisonnement d'un peu plus près.

Heil et Robb partent du constat selon lequel il semble difficile de concevoir une propriété qui soit purement catégorique. Pour eux, une telle propriété serait en quelque sorte épiphénoménale : elle existerait mais ne conférerait aucun pouvoir causal à l'objet qu'elle caractérise. Cela est problématique, et pas seulement eu égard au fait que postuler des entités épiphénoménales peut sembler vain. En effet, si l'on considère un exemple classique de propriété catégorique, comme être rond, il semble aller de soi qu'elle confère également un pouvoir causal particulier à l'objet qui l'instancie : par exemple, celui de rouler lorsque déposé sur un plan légèrement incliné. Cet exemple illustre bien le problème : il est difficile d'imaginer une propriété qui soit une caractéristique d'un objet mais qui ne renseigne pas sur le comportement qu'aurait cet objet dans certaines circonstances, donc qui ne lui confère aucun pouvoir causal. Pour Heil et Robb, il s'agit d'une contradiction dans les termes.



Puisque la thèse (D) affirme que toute propriété doit être soit dispositionnelle soit catégorique mais jamais les deux, si aucune propriété ne peut être uniquement catégorique alors nous devons affirmer que toutes les propriétés sont uniquement dispositionnelles, sans quoi nous devons abandonner (D). Cependant, pour Heil et Robb, l'idée d'une disposition « pure » mène à une régression à l'infini. Ils écrivent : « a disposition is *for* some manifestation with some reciprocal disposition partner. If every manifestation were itself purely dispositional, a regress looms, one which is characterizable only as being for something else, which is characterizable only as being for something else, which is characterizable only as... » (Heil et Robb 2003, p. 185). En l'absence de propriétés purement catégoriques et compte tenu de l'impossibilité de caractériser en quoi consiste une disposition pure, il ne reste qu'à abandonner l'idée que toutes les propriétés sont soit dispositionnelles soit catégoriques. C'est ainsi que Heil et Robb sont conduits à rejeter la distinction. Cela les pousse à préférer l'expression « propriétés qualitatives » pour désigner les propriétés catégoriques, « because [...] "categorical" has come to mean "non-dispositional" » (p. 185).

Nous n'entrerons pas plus en détails dans ces discussions sur la distinction entre propriétés dispositionnelles et catégoriques, mais il est important de retenir que les auteurs affirment que les tropes, qui constituent la réalité, sont à la fois des propriétés dispositionnelles et qualitatives. Mais cette réalité est-elle constituée uniquement de tropes ? Si oui, comment ces particuliers abstraits peuvent-ils constituer des objets concrets ? Nous allons maintenant nous tourner vers l'examen d'un concept qui divise les théoriciens des tropes : la substance.

### **2.1.3 Le concept de substance**

Les propriétés, entendues au sens de tropes, sont donc des particuliers abstraits qui constituent la réalité. Il s'agit là de la base de toute théorie des tropes. En revanche, il existe plusieurs versions de cette théorie, et un point sur lequel il est particulièrement important de les distinguer est la question de la substance. La notion de substance est un concept clef en ontologie car elle remplit un rôle explicatif de premier plan. En effet, lorsqu'on pense à une propriété, on a tendance à l'associer à un objet qui l'instancierait.<sup>43</sup> Il s'agit là d'un schéma classique en ontologie (qui fait bien sûr l'objet d'une multitude de variations et de débats) : la base de la réalité est composée d'une substance, qui possèdent différentes propriétés. Ce cadre théorique permet de rendre compte du changement *et* de la durée des objets qui peuplent le réel : lorsqu'un objet subit des changements, on peut les expliquer en termes de perte ou de gain de certaines propriétés. À l'inverse, même lorsqu'un objet subit de grands changements, on veut pouvoir affirmer qu'il reste le même en un certain sens,

---

<sup>43</sup> Et ce, que l'on considère les propriétés comme des universaux ou comme des particuliers. Dans le premier cas, on parle vraiment d'instanciation, et dans le second on peut parler de « possession », bien que plusieurs philosophes parlent tout de même d'instanciation dans le cas de particuliers (Heil 2003)

qu'il perdure. Enfin, la substance sert à expliquer la « nature » des objets concrets, ce qui fait leur particularité et qui permet d'expliquer leur individuation : le fait qu'on puisse les saisir, les reconnaître malgré des changements, et les associer à d'autres objets similaires. Dans un livre d'introduction aux débats contemporains en métaphysique paru en 2019, Anna Marmodoro et Erasmus Mayr résument ainsi les rôles explicatifs que remplit le concept de substance et que nous venons d'énumérer :

- i) Substances are bearers of properties, and of relations, which ontologically depend on them.
- ii) They underlie at least paradigmatic forms of qualitative change and are what remains the same during a change.
- iii) They are relatively stable insofar as they can exist for longer periods of time and can take on contrary properties.
- iv) They come as certain kinds of objects, having a nature that determines the kind of object they are and the conditions for their individuation at one time as well as for their diachronic persistence. (Marmodoro et Mayr 2019, p. 18)

C'est sur ce concept et sur la nécessité de lui faire une place au sein de leur ontologie que les théories des tropes divergent. Pour plusieurs philosophes, adopter une ontologie de tropes signifie tout simplement se débarrasser de la notion de substance. Dans un tel cas, il n'existe pas de *substance* qui posséderait des propriétés, les propriétés *sont* ce qui constitue les objets concrets qui peuplent la réalité. Ces positions font partie de la famille des théories du faisceau (*Bundle Theories*).<sup>44</sup> En effet, celles-ci considèrent que les objets qui composent le réel ne sont que des paquets ou des collections de tropes. Postuler une substance qui serait la base de tels objets est donc redondant.

L'alternative est de conserver un concept de substance distinct de celui de paquet ou de collections de tropes, et d'affirmer que les objets concrets qui peuplent le réel sont une combinaison d'un substrat de base et de propriétés possédées par celui-ci. Heil défend cette position dans son ouvrage de 2003, et ne considère pas que les objets concrets sont constitués de tropes. Pour lui, le concept de substance est nécessaire et il souhaite se distancer des philosophes qui adhèrent aux théories du faisceau. Il écrit :

Tropes have come to be identified with views championed by G. F. Stout, D. C. Williams, Keith Campbell, and Peter Simons, among others [...] These authors regard objects as "bundles" of tropes. This makes it appear that tropes are parts or ingredients of objects. Objects, however, are not made up of their properties in the way a clock is made up of its parts: screws, gears, a

---

<sup>44</sup> Les philosophes qui conçoivent les propriétés comme des universaux peuvent aussi rejeter le concept de substance et défendre que les objets ne sont ni plus ni moins que des ensembles (d'instanciations) de propriétés universelles. Toutefois, cette position est problématique, notamment concernant la possibilité de distinguer deux objets possédant les mêmes propriétés (Marmodoro et Mayr 2019, p. 29). Nous ne rentrerons pas ici dans les détails des critiques de cette position car cela nous écarterait de notre propos.

spring, an escapement, and a case. Parts of objects are objects, not properties. Properties—modes<sup>45</sup>—are particularized ways objects are. (Heil 2003, p. 128)

Nous revenons sur le terme « mode » un peu plus tard. Pour l'instant, il est important de comprendre que les théories des tropes se partagent en deux catégories, les théories du faisceau qui rejettent le concept de substance en considérant que les tropes remplissent toutes ses fonctions explicatives, et les partisans du concept de substance, comme Heil, qui voient le caractère *abstrait* des propriétés – rappelons que pour les adeptes des théories des tropes, les propriétés sont des particuliers abstraits – comme un élément distinguant ces dernières des objets *concrets* qu'elles caractérisent, ne pouvant donc pas en être les composants, au même titre que les pattes ou le dossier d'une chaise sont les composantes de celle-ci. Tenter de départager ces deux positions nécessiterait une présentation rigoureuse de leurs points forts et de leurs points faibles respectifs, une analyse que nous ne pouvons nous permettre ici. Nous reviendrons à la fin de la section 2.2 sur les différents termes utilisés pour désigner les propriétés lorsqu'elles sont conçues comme des particuliers abstraits, mais nous tâcherons de demeurer neutre sur la question des objets concrets : nous ne trancherons pas entre les théories du faisceau (qui les considèrent comme des paquets de tropes) et celles qui postulent deux types d'entités (un substrat et des propriétés lui appartenant ou le caractérisant).

Pour clore cette discussion qui visait à clarifier l'idée que les tropes sont des propriétés conçues comme des *particuliers abstraits* et pouvant être à la fois dispositionnelles et qualitatives, nous allons maintenant nous tourner vers une dernière distinction importante, celle entre les notions de prédicat et de propriété. En effet, certaines façons de concevoir cette distinction auraient conduit à une confusion concernant les propriétés, ce que dénoncent les théoriciens des tropes.

#### **2.1.4 Propriétés et prédicats : la théorie de l'image (*Picture Theory*)**

Il est important, afin de compléter ces remarques sur la notion de propriété, de la distinguer de celle de prédicat, avec laquelle elle est souvent associée et parfois confondue. Heil et Robb, nos deux auteurs principaux, discutent dans plusieurs de leurs textes d'une conception très répandue en philosophie bien que peu souvent formulée explicitement : la théorie de l'image. Il s'agit, très grossièrement, de l'idée selon laquelle chaque prédicat d'une langue désigne une propriété existant dans le monde. Une propriété est quelque chose qui existe, qui est instanciée par un objet à un instant donné et qui *n'est pas* une entité linguistique,

---

<sup>45</sup> Heil préfère le terme « mode ». Il choisit ce terme car il souhaite se distancer des philosophes qui voient les tropes comme des blocs de construction desquels les objets du réel seraient composés. Il y a des différences entre les positions de Heil et de Robb, certes. Bien que les deux auteurs aient cosigné un article dans lequel ils présentent une position commune (Heil et Robb 2003), il est évident dans d'autres de leurs écrits qu'ils divergent au moins sur un point crucial : le concept de substance, que Heil, contrairement à Robb, entérine. Le choix de termes ne semble toutefois pas très important et nous utiliserons toujours le mot « trope ».

contrairement à un prédicat. Par exemple, une boule de billard, un objet du monde, possède la propriété d'être ronde, quelque chose qui existe aussi dans le monde. Les prédicats, en revanche, sont des expressions linguistiques qui affirment quelque chose à propos des objets désignés par les termes singuliers<sup>46</sup>, de façon à dire quelque chose de vrai si les objets en question possèdent effectivement les propriétés exprimées par les prédicats. Ainsi, dans l'énoncé « Judith Butler est une philosophe », l'expression « Judith Butler » est un terme singulier : elle désigne un élément particulier du monde. L'expression « est une philosophe », elle, est un prédicat : elle attribue la propriété d'être une philosophe à Judith Butler. Cet énoncé est vrai puisque Judith Butler est effectivement une philosophe. Revenons à la théorie de l'image. Heil la définit ainsi dans son livre de 2003, *From an Ontological Point of View* : « language pictures reality in roughly the sense that we can “read off” features of reality from our ways of speaking about it » (p. 23).

Cette définition est très générale, et la théorie peut prendre de multiples formes. La plus forte d'entre elles stipule que *tout* prédicat va chercher une propriété existant dans le monde. D'autres versions, moins contraignantes, affirment que seulement certains prédicats réfèrent<sup>47</sup> à des propriétés (par exemple, les prédicats des langages scientifiques ou spécialisés). Selon Heil et Robb, cette théorie de l'image est implicitement présente chez beaucoup de philosophes, habituellement dans ses formes plus faibles (car les auteurs admettent que personne ou presque ne défend ou n'endosse la version forte). Cela engendre de nombreuses confusions dans les discussions concernant le statut de certaines propriétés, car au lieu de discuter directement celles-ci, il est question de prédicats. Une des caractéristiques communes aux théories ontologiques postulant des tropes est le rejet de cette conception.<sup>48</sup>

La raison qui pousse à adopter la théorie de l'image est le besoin de rendre compte de la signification d'un prédicat, et du fait qu'il puisse s'appliquer ou ne pas s'appliquer à un objet : en effet, il nous semble évident que le prédicat « est un être humain » s'applique à Noam Chomsky, alors qu'il ne s'applique pas à la Tour Eiffel. Une façon intuitive d'expliquer le sens d'un tel prédicat et notre capacité à déterminer s'il

---

<sup>46</sup> Les termes singuliers sont des expressions désignant des objets particuliers. En accord avec Frege, on peut les diviser en deux catégories : les noms propres et les descriptions définies.

<sup>47</sup> Ici, nous utilisons les expressions « référer », « désigner », « aller chercher » ou encore « s'appliquer à » pour éviter les répétitions et rendre la lecture moins aride. Nous n'aborderons pas les débats vivaces en philosophie du langage concernant la notion d'expression référentielle et la distinction entre sens et référence, car cela dépasse quelque peu notre propos. L'important est de concevoir les théories de l'image comme des assumptions implicites au sujet de la relation entre prédicats et propriétés, poussant à l'amalgame des deux dans les discussions à propos de l'existence des secondes.

<sup>48</sup> À notre connaissance, seule la philosophe Anna-Sofia Maurin est favorable à une ontologie de tropes tout en préconisant non pas l'abandon mais la refonte de la théorie de l'image afin de régler les problèmes auxquels cette dernière fait face. Maurin conçoit la théorie de l'image comme un cadre méta-ontologique et affirme qu'elle est nécessaire à la pertinence des travaux en ontologie. Elle écrit : « The Picture Theory fulfils two distinct, albeit related, functions. On the one hand, it describes the relationship that holds between language and reality. On the other, it, thereby, regulates ontological theorising [...] The Picture Theory must be replaced by a *new* Picture Theory » (Maurin 2008, p. 48).

s'applique ou non à un objet consiste à associer étroitement le prédicat en question avec la propriété qu'il désigne et qui est possédée par l'objet particulier que nous considérons, de même que par tous les objets auxquels ce même prédicat peut s'appliquer également. On voit s'esquisser ici la confusion profonde dont nous avons parlé plus haut entre deux sens du terme « propriété » : quelque chose d'unique, possédé par un objet, et quelque chose de commun, partagé par tous les objets tombant sous la coupe du même prédicat. Heil précise que ce qui pose problème est en fait un principe corolaire à la théorie de l'image, qu'il note  $\phi$  : « When a predicate applies truly to an object, it does so in virtue of designating a property possessed by that object and by every object to which the predicate truly applies (or would apply) » (Heil 2003, p. 26). Or, il y a deux idées distinctes ici, deux conceptions concurrentes de la notion de propriété. Les usages du terme « propriété » étant pluriels, et il ne faut donc pas les confondre entre eux ou les associer trop étroitement avec la notion de prédicat.

Dans cette section, nous avons présenté une conception particulière de la notion de propriété : les tropes, et clarifier quelques aspects de cette notion telle qu'elle est utilisée par les théoriciens qui nous occupent. Nous allons continuer à préciser cette notion et voir en quoi elle rompt avec une vision très largement répandue en philosophie : la vision en strates de la réalité.

## **2.2 Les niveaux de réalité**

Maintenant que la notion de propriété est un peu plus claire, il faut se pencher sur celle de niveau de réalité, afin de bien cerner où se situent les théories des tropes sur l'échiquier des positions ontologiques. En effet, toute la question du réductionnisme est intimement liée à l'idée que certains niveaux de la réalité sont plus fondamentaux – le niveau physique étant le plus fondamental – alors que d'autres le sont moins – il en est ainsi du mental – et sont éventuellement réductibles aux niveaux qui leur servent de base. Après avoir présenté cette vision, nous montrerons comment elle est liée à deux positions importantes en philosophie de l'esprit : l'identité des types et l'identité des occurrences. Cela va permettre de distinguer la théorie des tropes d'autres positions pouvant s'en approcher. Pour finir, nous verrons que la vision de strates de la réalité – vision faisant souvent office de posture par défaut – est rejetée par Heil et Robb, qui affirme qu'il faut lui préférer une ontologie moniste constituée uniquement de tropes.

### **2.2.1 Niveaux de description et niveaux de réalité**

La survenance, dont nous avons parlé au premier chapitre, est une relation entre des propriétés provenant de plusieurs niveaux de réalité hiérarchisés entre eux et dépendant tous d'un niveau fondamental, élémentaire. Dans nos discussions, par exemple, il était question de faire survenir les propriétés mentales sur des propriétés physiques, et ce afin de conserver à la fois l'efficacité causale et l'indépendance ontologique des premières tout en restant dans un paradigme physicaliste. Une telle relation suppose de concevoir les

différentes propriétés qui entretiennent une relation de survenance comme appartenant à différents niveaux, constituant la réalité. Cette vision du monde en strates (appelée *layered view of reality* dans la littérature), est abondamment discutée en philosophie analytique ainsi que dans d'autres disciplines (Heil 2003, p. 17, 18). Il faut toutefois distinguer l'idée de niveaux de *réalité* de celle de niveaux de *description*. Pour cela, notre clarification de la relation entre prédicats et propriétés va nous être utile. Puisqu'il est possible d'utiliser plusieurs prédicats appartenant à des champs lexicaux différents – voire à des théories différentes – pour décrire avec précision les mêmes événements, on dit que la réalité admet des niveaux d'*explication*, de *description*, ou d'*abstraction* différents. Cette idée, nous disent Heil et Robb, est bénigne. Il serait curieux de vouloir la remettre en question et probablement difficile de le faire. Comme exemple, on peut citer le cas classique de la température. Pour expliquer l'histoire causale d'un événement comme l'auto-inflammation d'une poche d'hydrogène, on peut avoir recours aux deux descriptions suivantes :

La poche d'hydrogène a atteint 400 degrés Celsius.

Les atomes d'hydrogène dans la poche ont atteint une énergie de 5595 Joules par mol.

Ces deux descriptions font le même travail explicatif, et décrivent le même événement, mais appartiennent à des niveaux d'abstraction différents. Cela ne semble pas poser de problème et s'applique à beaucoup d'autres domaines. Toutefois, toujours selon nos auteurs, une mauvaise compréhension de la distinction entre prédicats et propriétés pousse des philosophes à conclure, en constatant l'existence de ces différents niveaux de description ou d'abstraction, qu'ils correspondent à des niveaux de *réalité* différents. La popularité de cette vision en strates de la réalité en philosophie provient donc d'une part de la confusion entre prédicats et propriétés dont nous avons parlé plus tôt, et d'autre part – plus spécifiquement en philosophie de l'esprit – du fait de passer de l'impossibilité de réduire le champ lexical du mental à celui du physique à l'impossibilité de réduire les *propriétés* mentales aux *propriétés* physiques. Dans ce qui suit, Heil et Robb notent (P) l'idée selon laquelle les prédicats correspondent à des propriétés et appellent « semantic argument » cette inférence fautive qui va de l'existence de niveaux de description aux niveaux de réalité :

What has attracted philosophers to the idea that reality is layered? One likely source is (P) coupled with a semantic argument. Since the demise of logical behaviorism, few theorists have imagined that talk of the mental could be translated or analyzed into talk of the non-mental: mental and physical terms are semantically autonomous. No physical predicate, however complicated, means the same as "x feels pain." This is a point about our language (or concepts), not about the world. Some philosophers have nevertheless taken it to imply levels of reality (Heil et Robb 2003, p. 179-180)

Cette vision en strates fait face à beaucoup de problèmes. Parmi eux, celui de caractériser la relation qu'entretiennent les niveaux entre eux. Est-ce de la causalité? D'ordinaire, cette relation est considérée comme *diachronique*, faisant intervenir une certaine temporalité et liant des *événements* entre eux. Les

niveaux sont plutôt représentés comme entretenant une relation *synchronique*, simultanée. Quant aux relations synchroniques, il existe plusieurs façons de les concevoir. Est-ce de la réalisation? De la survenance? De la fondation (*grounding*)? Chacune de ces possibilités pose un certain nombre de problèmes, mais nous ne les explorerons pas tout de suite. Pour le moment, il est important de retenir que l'idée qu'il existe plusieurs niveaux de *description* est distincte de l'idée qu'il existe différents niveaux ou couches de *réalité*. Pour les adeptes des théories des tropes, la réalité ne possède pas plusieurs niveaux, correspondant à différents niveaux de description, mais un seul (pouvant être décrit de multiples façons). Nous allons revenir sur ce rejet de la vision en strates de la réalité un peu plus loin. Pour l'instant, nous allons nous servir de cette notion de niveau pour distinguer deux familles de solutions au problème de l'exclusion dont nous n'avons pas parlé au premier chapitre, et pour expliquer en quoi nous pouvons caractériser les théories des tropes comme appartenant à l'une d'elles.

### 2.2.2 Identité des types et identité des occurrences

Nous avons vu au premier chapitre que parmi les positions physicalistes, on trouvait les théories de l'identité qui jusqu'à l'avènement du fonctionnalisme étaient la famille de position la plus populaire. Toutefois, il serait faux de dire qu'elles sont tombées en désuétude après l'avènement du fonctionnalisme. Les partisans de la théorie de l'identité, encore très actifs dans les débats en philosophie de l'esprit, adhèrent tous au moins à la l'identité des occurrences, par exemple. Qui plus est, l'identité des types n'a pas dit son dernier mot. L'approche a d'ailleurs fait l'objet d'un ouvrage collectif paru en 2012 sous la direction de Simone Gozzano et Christopher S. Hill et intitulé *New Perspectives on Type Identity*. Pour saisir la particularité de la théorie des tropes, il est important de la distinguer de ces autres positions physicalistes. C'est ce que nous allons faire dans cette section.

Les théories de l'identité dont nous allons parler sont parfois nommées *psychophysical identity theses*. Elles peuvent se diviser en deux groupes, qui correspondent aux deux concepts que nous avons tirés de l'idée de propriétés dans la section 2.1 : l'identité des occurrences (appelée *token identity* ou *token physicalism*), d'une part, propose d'identifier les occurrences d'événements mentaux avec des occurrences d'événements cérébraux (les éléments qu'elle implique sont donc des particuliers). L'identité des types (*type-type identity* ou *type physicalism*), d'autre part, propose d'identifier les *propriétés* mentales (entendues donc au sens de types, ou d'universel, et non au sens de particulier ou de trope) avec des propriétés physiques.<sup>49</sup> La deuxième thèse est beaucoup plus forte, et implique la première. Toutefois, elle est également beaucoup plus difficile à défendre. Plusieurs adeptes de cette seconde thèse ont tenté d'en défendre la pertinence en

---

<sup>49</sup> Pour des discussions sur les rapports entre l'identité des types et l'identité des occurrences, voir l'ouvrage *New Perspective on Type Identities*, dirigé par Gozzano et Hill et paru en 2012.

mettant de l'avant la faiblesse de l'identité des occurrences. Kim lui-même, qui n'adhère pourtant pas exactement à l'identité des types, écrit dans un recueil de textes sur l'identité des types publié en 2012 : « In my view, one important reason why physicalism must be type physicalism, or at least concern itself with mental and physical types, is the dubious intelligibility of token physicalism as a distinctive physicalist doctrine » (Kim 2012, p. 167).

Pour Heil et Robb, toutefois, il n'est pas question d'adhérer à l'identité des types. La raison qui les pousse à rejeter cette thèse est à l'intersection des considérations que nous avons détaillées plus haut, et on commence ainsi à voir se dessiner la solution qu'une ontologie de tropes apporte au problème de l'exclusion : les propriétés mentales, en tant que particuliers abstraits, *sont* des propriétés physiques, mais les types mentaux ne sont pas réductibles à des types physiques. Les auteurs ne défendent donc pas une identité des types, mais une identité des occurrences, ou plus exactement, une identité des tropes :

In contending that mental properties are physical properties, we are not advancing a type-type identity theory. We regard it important to distinguish properties and types. We are happy to say that particular versions [...] of the Multiple Realizability and Semantic Arguments might establish that mental types are not physical types. But higher-level types are not ontological additions to our world; they are simply more abstract ways of characterizing physical properties. Our view is compatible with type-dualist views popular today. Puzzles about mental causation arise only if types and properties are identified. Is our view then just a token-token identity theory? Yes, although our version crucially amends familiar token-identity theories (Heil et Robb 2003, p.188)

Dans ce passage, les auteurs rejettent clairement l'identité des types, mais disent vouloir conserver une forme d'identité des occurrences. Les types ne sont que des catégories créées pour décrire le monde, mais ces catégories ne sont pas des catégories ontologiques : on retrouve bien notre distinction entre niveaux de description ou d'abstraction, et niveaux de réalité. Pour ce qui est des niveaux de description, aucun problème : Heil et Robb ne proposent pas de les réduire les uns aux autres. Toutefois, il ne faut pas en conclure qu'il est possible d'inférer, à partir de l'irréductibilité de ces niveaux de description, une irréductibilité de niveaux de réalité, correspondants à ces niveaux de description. Ici, c'est à nouveau l'amalgame entre prédicats et propriétés qui revient. Il ne peut être question de réduire ou non un niveau de réalité à un autre puisqu'il n'y a pas plusieurs niveaux de réalité.

Avant de terminer, nous allons distinguer différents termes qui peuvent être employés pour parler du concept ontologique de propriété en tant que particulier abstrait, afin de mieux saisir la distinction entre les différentes théories de l'identité. Trois termes nous intéressent particulièrement : « trope », bien sûr, mais également « occurrence » et « jeton » (qui sont tous deux employés parfois pour traduire « token »). Le lecteur pourra ainsi distinguer la théorie des tropes de l'identité des occurrences (la seconde peut sembler



assez proche de la première mais ne constitue pas exactement la même position), et avoir une meilleure idée de ce que le terme « trope » peut signifier pour les philosophes qui l'emploient.

Le terme « token », que nous avons rencontré dans notre discussion sur les théories de l'identité, est habituellement traduit par « occurrence » dans le cadre des discussions sur ces théories. Dans ce contexte, la notion d'occurrence se rattache davantage aux évènements, et les théories de l'identité des occurrences se contentent d'identifier toute occurrence d'un évènement mental avec une occurrence d'un évènement physique. Donald Davidson a été le premier à populariser cette thèse<sup>50</sup>. Pour lui, les évènements sont des particuliers indécomposables. Il refuse donc la conception kimienne des évènements présentée plus haut selon laquelle un évènement est l'instanciation d'une propriété par un objet à un moment du temps. Dans cette optique, l'acceptation de la théorie de l'identité des occurrences d'évènement n'oblige pas à identifier les propriétés mentales à des propriétés physiques pour pouvoir reconnaître qu'un évènement mental est identique à un évènement physique.

La plupart des théories des tropes sont cependant plus fortes que de simples théories de l'identité des occurrences. Pour voir pourquoi, il faut s'arrêter à la notion d'évènement que favorisent les défenseurs de la théorie des tropes. Pour Williams, les évènements sont eux-mêmes des tropes. Il écrit : « if a bit of perceptual behavior is a trope, so is any response to a stimulus, and so is the stimulus, and so therefore, more generally, is every effect and its cause [...] Causation is often said to relate events, and generally speaking any event is a trope: a smile, a sneeze, a scream, an election, a cold snap, a storm, a lightning flash, a conspiracy, perhaps a wave, and so on to such big and important events that they have proper names, like the Passover and Lulu the H-bomb explosion » (Williams 1953b, p. 172). Une théorie qui accepte l'idée qu'un évènement est un trope, c'est-à-dire un particulier abstrait, implique une identité des occurrences d'évènement mais va plus loin, en identifiant les propriétés mentales à des propriétés physiques, ou plus exactement en clarifiant ce qu'on entend par « propriété » et en reléguant la question du physique ou du mental aux classes de ressemblance. Cette position est celle de Heil et de Robb dans les textes que nous discutons dans ce mémoire. Les deux auteurs la choisissent car ils tentent de répondre au problème de l'efficacité causale du mental formulé par Kim dans ses propres termes, soit en faisant valoir que les propriétés mentales ont des pouvoirs causaux.

Il est important de saisir la différence entre une simple théorie de l'identité des occurrences et la théorie des tropes défendue par Heil et Robb, car elle permet de saisir la spécificité de cette dernière. Celle-ci repose sur le rejet de la vision en strates de la réalité dont nous avons parlé un peu plus tôt. La différence

---

<sup>50</sup> Cf. Davidson, 1969.

entre une simple identité des occurrences et la théorie des tropes est la compatibilité avec cette vision en strates. L'identité des occurrences est tout à fait compatible avec une telle vision : l'identité étant restreinte aux évènements, on peut défendre à la fois une identité des occurrences et un dualisme des propriétés. On peut par exemple défendre un dualisme des types, ou même une forme de physicalisme non réductionniste dans lequel des propriétés mentales particulières surviennent sur des propriétés physiques tout en étant irréductibles à celles-ci. Or, la théorie des tropes est résolument réductionniste, et rejette également, on l'a vu un peu plus tôt, le dualisme des types : il n'existe pas de niveaux de réalité, seulement des niveaux d'abstraction – les classes de ressemblance – qui dépendent de nos grilles de lecture.

Le terme « token » peut aussi se traduire en français par le terme « jeton ». Dans le cadre de la théorie des tropes, ce terme nous paraît pertinent, car il permet une compréhension intuitive de ce qu'est un trope : un « jeton » parmi d'autres. Pour illustrer cela – et il ne s'agit là que d'une analogie faite pour illustrer la notion – il suffit de penser à un jeu de société dans lequel il y a beaucoup de pièces distinctes : des pions, des dés, des figurines en tout genre. Souvent, on trouve plusieurs jetons de la même sorte dans un jeu, petites pièces qui ne sont que cela : des jetons (de tout petits cylindres servant à compter les points, par exemple). Ces derniers se ressemblent beaucoup, mais ne sont évidemment pas identiques dans toutes leurs caractéristiques. Ils présentent des différences (de légères variations de taille, de masse, etc.) qui, bien que non perceptibles au premier regard et non pertinentes pour l'usage qui en est fait, sont bien réelles. Les jetons du jeu sont un peu comme les tropes : ils tombent sous des classes de ressemblance (les jetons bleus, les jetons rouges, les jetons triangulaires, les jetons en plastique, etc.), mais diffèrent tous sous certains aspects, et sont donc tous uniques.

Bien que le terme « jeton » ait son intérêt, nous utiliserons seulement le terme « trope » dans le reste de ce mémoire car il est plus répandu.

### **2.2.3 La vision en strates rejetée**

Tel que mentionné plus haut, les auteurs vont plus loin que la simple dissociation entre les affirmations au sujet des niveaux de description et celles au sujet des niveaux de réalité. Pour eux, il faut purement et simplement rejeter la vision en strates de la réalité. Les propriétés sont tout ce qui constitue le réel, et celui-ci n'est pas étagé, il n'est composé que d'une seule sorte d'entité : les tropes.

Reality has but one level—or, if the idea that reality has one level encourages the idea that it might have more, there are no levels of reality. We do not need levels to be realists about states of mind, trees, statues, and people. We do not need a commitment to ontological levels to accommodate irreducible, projectable predicates definitive of everyday domains and those of the special sciences. We may find it occasionally useful to speak of levels of description or

explanation, but these must not be confused with levels of being or promote the image of a layered world. (Heil 2003, p. 50)

Le fait que ces tropes et les événements qu'ils contribuent à générer tombent sous la coupe de différents prédicats et admettent différents niveaux de description ne doit donc pas pousser à considérer qu'ils puissent être classés selon différentes catégories ontologiques mutuellement exclusives correspondant à ces différents prédicats. De telles affirmations, en effet, obligent ensuite à postuler des relations complexes de réalisation ou de survenance entre ces différents niveaux afin de concilier les différentes intuitions et thèses que nous avons discuté au premier chapitre. La caractérisation des relations entre les niveaux est la raison principale qui pousse nos auteurs à proposer l'abandon de la vision en strates de la réalité. Comme nous l'avons mentionné plus haut, postuler de tels niveaux pose toutes sortes de problèmes théoriques auxquels il faut s'atteler pour justifier l'impact que des propriétés de niveaux supérieurs pourraient avoir sur le monde physique, et ce afin de ne pas tomber dans l'épiphénoménisme. Rappelons-nous une partie cruciale du puzzle présenté au premier chapitre : comment respecter à la fois l'indépendance ontologique ET l'efficacité causale d'une propriété qui n'appartiendrait pas au « niveau physique », dans lequel les relations de causalité sont pourtant sensées prendre place ? Pour Heil et Robb, et les adeptes des ontologies de tropes en général, il est préférable de laisser de côté cette conception et d'abandonner avec elle la théorie de l'image, qui en est la source.

Les théories des tropes semblent donc être des monismes, et ne traitent pas de la notion de propriété comprise comme élément commun à plusieurs objets comme d'une catégorie ontologique. Toutefois, il faut tout de même rendre compte de cette notion – que Heil et Robb désignent par l'expression « classe de ressemblance ». Avant de conclure ce chapitre, nous allons la préciser quelque peu et voir en quoi on peut l'analyser à l'aide d'une perspective fonctionnaliste.

### **2.3 Classe de ressemblance et rôle fonctionnel**

Dans cette section, il est question d'une tentative de conciliation d'une ontologie de tropes avec une approche fonctionnaliste de la notion de propriété. L'article sur lequel nous nous basons est celui d'Ann Whittle, « A Functionalist Theory of Properties », dans lequel elle affirme que la théorie du rôle fonctionnel – une des deux variantes du fonctionnalisme que nous avons présentées au premier chapitre – peut être couplée à une ontologie de tropes pour donner une théorie générale des *propriétés*. Dans un premier temps, nous allons expliquer pourquoi Whittle préfère la théorie du rôle fonctionnel à celle de la spécification fonctionnelle pour accompagner une ontologie de tropes en présentant une méthode employée par les fonctionnalistes pour rendre compte des propriétés qu'ils discutent : la méthode Ramsey-Lewis. Puis, nous allons expliquer en quoi elle considère non seulement qu'une ontologie de tropes et une analyse

fonctionnaliste de la notion de propriété sont compatibles, mais également que ces deux positions sont mutuellement bénéfiques. En effet, chacune règle des problèmes théoriques rencontrés par l'autre.

### **2.3.1 La technique Ramsey-Lewis et le projet d'analyse fonctionnelle de la notion de propriété**

L'idée centrale du fonctionnalisme est de concevoir les états mentaux comme des états fonctionnels, et d'ainsi expliquer leur efficacité causale sur le monde tout en conservant l'indépendance du mental par rapport au physique – l'indépendance des types mentaux et physiques, à tout le moins. Ce que Whittle appelle la technique Ramsey-Lewis a été développée par David Lewis pendant ses travaux sur la théorie de l'identité. En s'inspirant des écrits du mathématicien Frank Ramsey et des travaux du philosophe Rudolf Carnap sur ces derniers, Lewis a cherché à offrir un outil d'analyse fonctionnelle de la notion de prédicat mental. Elle apparaît chez celui-ci dès 1966 avec son article « An Argument for the Identity Theory », et est développée fréquemment par la suite dans ses travaux en philosophie de l'esprit (elle est particulièrement explicite dans « How To Define Theoretical Terms », paru en 1970 et dans « Psychophysical and Theoretical Identifications », paru en 1972). Cette technique permet d'offrir une définition fonctionnelle d'un prédicat de n'importe quel niveau de description, en autant que l'entité qui satisfait celui-ci remplit un rôle fonctionnel. Elle est surtout utilisée, dans la littérature, pour défendre la position fonctionnaliste en philosophie de l'esprit, et son but est donc d'offrir des définitions fonctionnelles des prédicats mentaux.

La technique Ramsey-Lewis consiste à reformuler une théorie employant des termes pouvant être imprécis ou problématiques, mais dont l'étude des référents nous importe (on voit ici le cas évident des termes mentaux), à l'aide d'autres termes d'une théorie employant des termes plus précis, mieux définis, et dont les référents peuvent être étudiés de manière efficace. L'idée est de systématiser la relation qu'entretiennent les référents des termes de ces deux théories, afin d'échapper aux problèmes rencontrés par la première.

Nous allons voir qu'une approche fonctionnaliste qui utilise cette technique de formalisation de la notion de propriété peut aider la théorie des tropes : elle précise sa notion de classe de ressemblance en lui donnant une définition plus précise. Toutefois, il nous faut également aborder la raison qui peut pousser à entériner la théorie des tropes en plus de cette conception fonctionnelle de la notion de propriété, afin de justifier le recours à une telle ontologie. Quel est l'intérêt pour un fonctionnaliste d'adopter une telle position ? Pour Whittle, ce n'est qu'à la condition d'accepter la théorie des tropes qu'il est effectivement possible de produire une analyse pertinente de la notion de propriété. Le projet d'analyse fonctionnelle de la notion de propriété est entre autres celui de Sydney Shoemaker, qui affirme dans un article de 1981 que la technique Ramsey-Lewis permet de faire de toute propriété une propriété fonctionnelle (Shoemaker 1981, p. 275). Cette question de la nature des propriétés est un débat à la fois ancien et vivace en philosophie, et on comprend

donc aisément l'intérêt que Whittle et bien d'autres portent à cette proposition de Shoemaker. Voyons quelle est l'idée de ce dernier.

La théorie du rôle fonctionnel, nous l'avons vu, affirme qu'avoir une propriété fonctionnelle (une propriété mentale, par exemple), consiste à être dans un certain état (cérébral, dans le cas des humains) qui joue un rôle causal particulier. Les propriétés fonctionnelles sont ainsi conçues comme des propriétés relationnelles d'ordre supérieur. Avoir une propriété fonctionnelle consiste à avoir la propriété d'être dans un certain état qui joue tel ou tel rôle dans telle ou telle circonstance. Les propriétés fonctionnelles sont donc un type d'entités distinctes des états qui les réalisent. Cela signifie que tout ce qu'il y a à dire des propriétés fonctionnelles est lié au rôle causal que remplissent d'autres types d'entités. Comme le note Whittle : « Their whole nature is exhausted by facts about what other entities will do, given certain circumstances » (Whittle 2008, p. 68). Ainsi il est nécessaire, pour qu'une propriété existe, que quelque chose remplisse le rôle qui la définit. Jusque-là, tout semble bien parti. Toutefois, si l'on tente d'appliquer cette définition à toutes les propriétés, alors on fait face à un dilemme : accepter une hiérarchie infinie de propriétés fonctionnelles réalisant les rôles fonctionnels les unes des autres, ou bien postuler de nouvelles entités qui peuvent remplir le rôle qui définit les propriétés. Whittle ne rejette pas a priori l'option d'une hiérarchie infinie de propriétés, mais elle affirme qu'il s'agit d'une question empirique, qui doit être réglée par les sciences : « the issue of how many properties there are seems to be an empirical one. If our scientific theories inform us that there are an infinite number of properties, fine » (p. 70). La possibilité n'est pas exclue, nous ne pouvons présumer qu'il y a un niveau fondamental à la réalité – le niveau physique – et que la science du futur ne nous amènera pas toujours plus loin dans l'infiniment petit. Le problème se complique cependant lorsque l'on cherche, comme Shoemaker, à appliquer l'analyse fonctionnelle à toutes les propriétés et non pas seulement aux propriétés mentales. Si toutes les propriétés sont des propriétés fonctionnelles, avoir n'importe quelle propriété consistera à avoir une autre propriété qui remplit un certain rôle causal, ad infinitum. Il ne restera plus aucun état pour occuper les rôles fonctionnels. Heureusement, il existe une autre façon d'aborder la question qui peut prévenir cette régression : avoir recours à des entités n'étant pas des propriétés fonctionnelles, mais remplissant les rôles définissant ces dernières. Ces entités, ce sont les tropes.

### **2.3.2 Théorie du rôle fonctionnel et ontologie**

Voyons si une ontologie de tropes est compatible avec une analyse fonctionnaliste des propriétés. Dans son article, Whittle affirme que la tentative d'analyse fonctionnelle de la notion de propriété de Shoemaker et une ontologie de tropes sont deux positions non seulement compatibles mais présentant des avantages réciproques. Pour elle, cette combinaison résout l'un des problèmes de chaque position. Pour la

théorie des tropes, l'avantage est de clarifier la notion de classe de ressemblance, dont nous avons dit un mot au chapitre précédent. Pour la théorie fonctionnaliste, l'avantage est d'éviter de sombrer dans une hiérarchie infinie de propriétés (l'un des casse-têtes de la théorie du rôle fonctionnel). Nous allons revenir sur les détails de ces deux positions en précisant ces deux avantages mutuels que Whittle met de l'avant.

Dans le chapitre précédent, nous avons défini brièvement la notion de classe de ressemblance, mais nous avons mentionné qu'elle était parfois critiquée comme manquant de précision. Cette notion peut en effet paraître floue, et les critères requis pour classer un trope dans l'une ou l'autre des classes n'étaient pas suffisamment bien définis. Toutefois, à l'aide d'une analyse fonctionnelle, Whittle affirme que cette ambiguïté est levée :

the analysis aids the cause of trope theory, by presenting us with a conception of what resemblance between tropes involves. The functional analysis offers a way of grouping tropes into sets which stand for natural properties. The set of tropes which is the property F-ness, on this view, consists of all those tropes which realise the functional role specified in the Ramsey-Lewis definition for property F-ness (Whittle 2008, p. 79)

Ici, Whittle utilise le terme « propriétés » pour parler d'universaux. Bien qu'elle demeure neutre dans l'article quant à l'existence d'universaux, l'idée est surtout de distinguer les usages des termes « propriété » et « trope » (elle dit d'ailleurs explicitement un peu plus tôt dans le texte : « properties or universals are logical constructions out of tropes »), et de se pencher sur la relation entre les deux. Pour elle, une classe de ressemblance de tropes est un ensemble de tropes remplissant le rôle fonctionnel d'une propriété tel que formulé par la technique Ramsey-Lewis. Prenons la douleur comme exemple. Considérons trois tropes : la douleur d'un humain, la douleur d'un chat et celle d'un Martien. Si ces trois tropes remplissent la même fonction, le même rôle causal, interprétable à l'aide de la technique Ramsey-Lewis, alors ils appartiennent à la même classe de ressemblance : celle des douleurs. La notion de classe de ressemblance se dote ainsi d'un critère plus précis, et est donc beaucoup plus claire.

De plus, elle fait le lien entre les tropes et les niveaux de description avec lesquels on les décrit. Whittle pense donc qu'une analyse fonctionnaliste de la notion de propriété aide à mieux définir la notion de ressemblance, nécessaire à une ontologie de tropes.

En ce qui concerne la théorie du rôle fonctionnel, Whittle affirme qu'une ontologie de tropes vient l'aider en clarifiant ce qu'elle appelle l'affirmation insaisissable de Shoemaker (« Shoemaker's elusive claim ») selon laquelle les propriétés peuvent être considérées comme des pouvoirs causaux (Whittle 2008, p.80). Que veut-elle dire ? Une des contributions philosophiques principales de Shoemaker est sa théorie causale des propriétés, selon laquelle les propriétés ne sont que des pouvoirs causaux. Cette position procède d'une conception exclusivement dispositionnelle des propriétés et refuse l'existence de propriétés catégoriques, qui

sont plus que de simples pouvoirs causaux et qui peuvent être caractérisées par des éléments intrinsèques leur appartenant. Nous avons introduit la distinction entre ces deux notions de propriétés dans la section 2.1.2. Puisque, comme nous l'avons mentionné dans la section en question, il n'est nul besoin pour un théoricien des tropes de trancher dans ce débat, la position de Shoemaker est compatible au moins a priori avec une ontologie de tropes. Or, Whittle affirme qu'une ontologie de tropes serait en fait nécessaire à celle-ci. Nous avons parlé dans la section précédente du risque de hiérarchie infinie de propriétés fonctionnelles réalisant les rôles fonctionnels les unes des autres ; c'est précisément là qu'une ontologie de tropes intervient. Rappelons-nous : l'idée n'est pas qu'une telle hiérarchie est problématique en elle-même, ou qu'elle est une impossibilité théorique. Le problème est lié à l'objectif de départ du projet d'analyse fonctionnelle, qui est de donner une analyse du *concept* de propriété, et il faut donc qu'une telle analyse porte sur *toutes* les propriétés. Or, se contenter de faire une hiérarchie infinie de propriétés se définissant les unes les autres n'est pas satisfaisant. Comme le note Whittle : « in order to achieve a functional role analysis of all properties, we need to show how the type of entity in question, namely the properties, can be analysed as abstractions from a different kind of entity » (Whittle 2008, p. 72). Les tropes sont ces entités distinctes des propriétés fonctionnelles dont ils réalisent le rôle causal, entités qui ne sont pas elles-mêmes caractérisées de manière fonctionnelle, ce qui nous évite la régression à l'infini. Il nous faut donc, pour pouvoir offrir une définition fonctionnelle des propriétés et être capable de l'étendre à *toutes* les propriétés, trouver des entités réalisant ces propriétés tout en étant indépendantes. Les tropes, en tant que particuliers abstraits, peuvent remplir ce rôle. Pour Whittle, il est clair que l'on peut analyser toute propriété en termes d'ensemble de tropes remplissant un rôle fonctionnel identique, tel que décrit par sa définition Ramsey-Lewis, il s'agirait même de la seule analyse satisfaisante de la notion de propriété (p. 74). Cette notion recoupe celle de classes de ressemblance qu'on a déjà croisée dans les théories de Heil et de Robb. Nous proposons donc de concevoir les classes de ressemblance comme des ensemble de tropes remplissant le même rôle fonctionnel.

## 2.4 Conclusion

Dans ce chapitre, nous nous sommes intéressés à la position selon laquelle la réalité est composée de tropes, c'est-à-dire de propriétés particulières, dont certaines se ressemblent plus ou moins les unes les autres, et peuvent être classées dans des catégories de notre création, selon leurs ressemblances subjectives (de notre point de vue). Nous avons également vu comment la théorie des tropes pouvait s'accompagner d'une conception fonctionnaliste de la notion de propriété (conçue comme élément commun à deux objets distincts), afin de clarifier ces classifications. Comme nous allons le voir, la théorie des tropes permet de résoudre le problème de l'exclusion présenté au premier chapitre en proposant une identité des propriétés mentales et des propriétés physiques : une propriété mentale et ce que certains auteurs considèrent comme

sa base de réalisation ne forment en fait qu'un seul et même trope, que nous associons à plusieurs classes de ressemblance différentes en fonction de nos critères descriptifs et théoriques (mental, physique, etc.).



## Chapitre 3 : Succès et limites de la théorie

### 3.0 Introduction

Dans ce chapitre, nous allons nous intéresser aux succès explicatifs de la théorie des tropes, à sa compatibilité avec d'autres théories philosophiques, ainsi qu'à la solution qu'elle offre au problème de l'exclusion causale. Nous allons d'abord chercher à montrer que la théorie des tropes, en tant que position ontologique, est pertinente et légitime. Nous montrerons qu'elle peut donner des réponses à des puzzles traditionnels en métaphysique, contrairement à ce que font valoir certaines critiques (3.1), et nous verrons qu'elle est acceptable parce que l'antiréalisme auquel elle conduit n'est pas problématique pour la science (3.2). Ensuite, nous reviendrons sur le problème de l'exclusion. Nous verrons d'abord en quoi la théorie des tropes y apporte une réponse réductionniste (3.3), puis en quoi il est intéressant pour les réductionnistes d'y adhérer (3.4). Enfin, nous aborderons une critique de cette utilisation de la théorie des tropes pour répondre au problème de l'exclusion (3.5).

### 3.1 Trois critiques métaphysiques de la notion de tropes

Les critiques de la théorie des tropes qu'on trouve dans la littérature peuvent être divisées en deux catégories. La première regroupe celles qui portent sur l'application de la théorie à une question philosophique particulière – nous tenterons de répondre à celle portant sur l'efficacité causale du mental dans la dernière section. La deuxième regroupe celles qui portent sur les difficultés théoriques de la position en elle-même. Dans cette section, nous allons présenter et tenter de répondre à trois de ces critiques. Celles-ci se retrouvent chez plusieurs philosophes, mais nous allons nous concentrer sur la formulation qu'en font Anna Marmodoro et Erasmus Mayr dans leur ouvrage *Metaphysics : An Introduction to Contemporary Debates and Their History*, paru en 2019. Nous concluons que la théorie des tropes est une position ontologique acceptable, car elle répond à ces différents puzzles traditionnels en métaphysiques.

La première de ces critiques affirme qu'il est impossible pour la théorie des tropes de rendre compte de la notion de ressemblance. Selon Marmodoro et Mayr : « the most obvious problem with this variant of the bundle theory is how to explain the similarity between different tropes and between bundles [...] we couldn't say that this page of the book has the same color as another one, since each page has its specific particular color that is different from the color of every other page » (p. 31). L'idée en arrière de la critique est que si la notion de propriété commune – la couleur partagée par les deux pages dans leur exemple – refait surface, alors il est superflu de parler de trope, puisqu'une des motivations de la position est de ne pas avoir recours aux universaux pour décrire la réalité. Les théories des tropes doivent donc expliquer la ressemblance entre plusieurs objets sans recourir à une notion comme celle d'universel.

Mais la critique ne s'attaque pas directement à la notion utilisée par Robb, Heil ou Whittle. En effet, les théoriciens des tropes proposent d'expliquer la ressemblance entre deux objets à l'aide des classes de ressemblance, qui ne sont pas des universaux instanciés par des propriétés particulières que seraient les tropes. Ce ne sont pas les tropes qui sont des instances de propriétés, ce sont les classes de ressemblance qui sont des ensembles de tropes. Il y a, on peut le rappeler, certaines théories des tropes qui prétendent être neutres sur la question de l'existence des universaux.<sup>51</sup> Ce sont cependant les classes de ressemblance qui peuvent rendre compte de ce que nous percevons comme des similarités entre des objets. C'est d'ailleurs sur la base de ces constats de ressemblance que nous créons, changeons ou abandonnons les classes en question. Si l'idée de la critique de Marmodoro et Mayr est qu'il faut rendre compte de l'*exacte* ressemblance de certaines propriétés (au sens d'universel) que posséderaient des objets ou des faisceaux, la réponse de la théorie des tropes serait que les propriétés des objets ou faisceaux ne sont jamais complètement identiques. Elles peuvent se ressembler, auquel cas certains des tropes qui caractérisent un faisceau ou un objet appartiennent probablement aux mêmes classes de ressemblance, mais elles ne peuvent être parfaitement identiques, car les objets ou faisceaux possèdent des tropes différents. Pour ce qui est de la façon de rendre compte de la ressemblance entre les tropes eux-mêmes, cela dépendra de la théorie retenue. Selon celle de Whittle, par exemple, deux tropes appartiennent à la même classe de ressemblance s'ils remplissent le même rôle causal. La ressemblance ne peut néanmoins pas aller plus loin car aucun trope n'est identique à un autre, ils sont des particuliers au sens fort, et chaque trope est unique.

La seconde critique que Marmodoro et Mayr adressent aux théories des tropes concerne spécifiquement les versions de celles-ci qui rejettent le concept de substance – les théories du faisceau, dont nous avons parlé au deuxième chapitre. Selon les philosophes, elles mèneraient à une impasse car elles ne pourraient garantir la distinction entre deux amas de tropes « identiques » que si l'on abandonne la motivation originelle de la théorie :

the main aim of the bundle theory, we saw, was to account for substances entirely in terms of properties. The particularity of substances was meant to be "explained away" in terms of coinstantiated properties. Now it seems we are able to salvage the bundle theory from the objection that two identical bundles would collapse into one, only if we understand properties in some way as particulars. For in order to make the bundles of properties behave more like the particulars that we take objects to be, we had to make our properties look like particulars (Marmodoro et Mayr 2019, p. 31-32)

Il est important de noter que les théories du faisceau – les théories ontologiques qui rejettent le concept de substance – se divisent en deux catégories : les théories des tropes d'une part, et les théories qui

---

<sup>51</sup> On pourrait d'ailleurs postuler des universaux qui regrouperaient les tropes en des classes de ressemblance « parfaites », mais qui seraient inconnues (et peut-être impossible à connaître ?) des agents tentant de les classer.

font de ces amas de particuliers des amas d'instanciations d'universaux. Pour ces théories, les universaux sont bien réels, et tout ce qui existe est une instanciation d'un ou plusieurs de ces universaux. Le début de la critique de Marmodoro et Mayr porte sur ce type de théories du faisceau, dans lesquelles les propriétés sont des catégories universelles. Selon ces théories, le concept de substance n'est pas nécessaire puisque tout est constitué de groupes d'instanciations de propriétés universelles. L'objection qui leur est opposée est que si elles ne veulent pas conserver le concept de substance, elles doivent rendre compte de l'aspect particulier des objets du monde qui possèdent ces propriétés. La question qui se pose alors à ces théories est celle de savoir comment elles peuvent distinguer deux entités possédant les mêmes propriétés.

Nous ne discuterons pas les réponses que peuvent apporter les théories du faisceau ayant recours aux universaux, mais nous allons dire quelques mots de la façon dont les théories des tropes peuvent y répondre. En effet, celles-ci aussi doivent pouvoir apporter une réponse à cette critique. Les tropes sont des particuliers, et par hypothèse, ils sont uniques. Mais une théorie des universaux pourrait tout aussi bien affirmer que chaque instanciation d'un universel est unique. Cela ne suffit pas à expliquer ce qui distingue deux objets instanciant exactement les mêmes universaux. La question est sensiblement la même pour les théories des tropes : comment distinguer deux amas de tropes dont les tropes appartiennent exactement aux mêmes classes de ressemblance ? Face à ce problème, une théorie du faisceau utilisant des tropes et non des universaux peut sembler une meilleure option : il semble difficile de défendre que les différentes instanciations du même universel sont uniques et diffèrent suffisamment pour permettre de distinguer deux objets les possédant. Après tout, une des fonctions principales de ce type de théories est d'expliquer efficacement la *ressemblance* entre deux entités – elles instancient la même propriété. Pour rendre compte de la *différence* entre deux instanciations de celle-ci, il faudrait par exemple en appeler à l'individuation spatiotemporelle (deux instanciations du même universel diffèrent si et seulement si elles n'ont pas les mêmes coordonnées spatiotemporelles), ce qui pourrait soulever d'autres critiques. Les tropes, en revanche, ne sont pas des instanciations d'universaux. Ce sont des éléments uniques, individuels, constituant la réalité. Ils sont d'emblée conçus comme des entités particulières. Les faisceaux de tropes sont donc, *ipso facto*, uniques eux aussi. Leur ressemblance, nous l'avons mentionné plusieurs fois, est contingente et est fonction d'observations extérieures. Cette réponse au problème de l'individuation (qui est un problème classique des théories du faisceau) est parfois appelée individuation primitiviste. C'est la principale position sur la question que les philosophes qui défendent une théorie des tropes entérinent.

La troisième critique concerne elle aussi les théories du faisceau, mais s'adresse spécifiquement aux théories des tropes. Elle affirme que ces dernières ne peuvent pas rendre compte de la possibilité pour un objet qui perd ou gagne une propriété de demeurer le même dans le temps. Ce rôle étant traditionnellement

rempli par le concept de substance, les théories des tropes qui s'en passent ne seraient pas en mesure d'expliquer cette persistance temporelle des objets en dépit de changements. Marmodoro et Mayr écrivent :

If a substance were just a collection of properties, then strictly speaking it couldn't survive any change. If one property was lost and another gained, we would have a different collection of properties, on the assumption that what makes a collection the same thing at different times is that it is composed of the same component items. Consequently, two collections are different if the things collected within them are different. However, the particulars that interest us change (qualitatively) all the time, while remaining (numerically) the same. The point will become intuitively clear if we consider a human being as an exemplary substance. A human being changes radically from childhood to adulthood. How could a person be just a collection of properties when her/ his properties change all the time? (p. 32)

Pour les auteurs, il est important d'expliquer comment les objets ou les faisceaux persistent à travers le temps, car ils doivent être capables d'endurer des changements. Un deuxième exemple, similaire à celui de l'être humain qui grandit est développé par les auteurs un peu plus loin : celui du bateau de Thésée. Au fil du temps, le bateau se serait dégradé progressivement, et on en aurait changé les planches de bois une à une. Après quelques années, il ne reste plus une seule des planches d'origine. Pourtant, on aurait conservé le bateau, tel un monument sur une place de la cité, une fois celui-ci revenu de son périple. La question est alors : au bout de plusieurs années, lorsque toutes les pièces du bateau ont été remplacées, peut-on dire qu'il s'agit du bateau de Thésée ? S'agit-il d'un autre bateau, d'une copie du premier, y-a-t-il rupture au moment où l'on change la première planche ? Au moment où l'on change la dernière pièce d'origine ? Y a-t-il continuité même lorsque plus un seul élément du navire d'origine n'est présent sur la place ? Regardons de plus près comment la théorie des tropes traite ce cas ainsi que celui des êtres humains qui changent au fil du temps.

Commençons par le cas d'un objet concret qui n'est pas un être vivant, celui du bateau de Thésée. Ce que nous disent Marmodoro et Mayr, c'est que la théorie des tropes ne serait pas en mesure d'expliquer en quoi le bateau de Thésée demeure le même lorsqu'on remplace certaines de ses pièces, car alors, sans le concept de substance – un élément qui demeurerait et ce peu importe les changements subis par l'objet – il ne s'agirait plus du même objet, et ce dès la perte (ou le gain) du premier trope. Notre réponse, c'est qu'il n'est pas dans l'idée d'une telle théorie d'affirmer qu'un objet demeure le même dans un sens primordial (comme c'est le cas dans le cadre d'une théorie selon laquelle un objet possède une substance qui demeure radicalement inchangée quoi qu'il arrive à l'objet) lorsqu'il perd ou gagne un trope. En fait, la lecture de la situation se ferait dans l'autre sens : lorsqu'un faisceau subit des changements de tropes, il n'est plus le même faisceau. On peut alors imaginer deux cas de figure. Ou bien ces changements sont imperceptibles pour des observateurs tels que nous, auquel cas nous continuerons à parler du même objet avant et après les changements, sans même nous rendre compte que les changements ont eu lieu. Ou bien, comme dans le cas du bateau de Thésée, nous pouvons percevoir ces changements, et même en parler, en discuter. Certains

pourront alors considérer qu'il s'agit du même objet, et d'autres que celui-ci n'est plus ce qu'il était. Les jugements vont d'ailleurs différer grandement d'une personne à l'autre : si l'on sondait Athènes régulièrement, il est probable que l'opinion populaire évolue au fur et à mesure de l'entretien du navire et du remplacement de ses pièces. Pour certains, remplacer ne serait-ce qu'un clou dénaturerait ce vaisseau sacré. Pour d'autre, la voile, la proue, ou encore le gouvernail pourraient être l'élément principal, l'âme du bâtiment. Ceux-ci reconnaîtraient alors le symbole de leur héros jusqu'à ce que cette pièce clef soit usée par le temps et soit remplacée. Toutefois, ces jugements a posteriori n'ont aucune importance eu égard à la réalité : le faisceau n'est plus le même dès lors qu'un seul changement a lieu. Celui-ci peut très bien conserver certains des tropes qu'il avait, mais s'il perd ou gagne un trope, il n'est plus le même faisceau. En fait, il semble plutôt curieux d'affirmer qu'un faisceau qui perd ou gagne un trope est le même faisceau : il faudrait faire appel à un concept comme celui de substance, que les théories du faisceau rejettent. Selon cette lecture, il n'est pas problématique d'abandonner l'idée de rendre compte de la persistance ontologique d'un objet malgré les changements qu'il subit. La persistance est au niveau de notre perception des objets concrets, et c'est pourquoi des exemples comme le bateau de Thésée ne feront pas l'unanimité : nous ne percevons pas toutes et tous les changements de la même façon, et ne jugeront pas qu'un objet est « dénaturé » selon les mêmes critères.

Pour ce qui est de l'exemple de l'être humain, la même logique peut être appliquée. Le cas semble plus délicat de prime abord car on peut être plus enclins à considérer qu'un humain demeure le même dans un sens primordial entre ses 2 et ses 82 ans qu'à considérer que le bateau de Thésée est encore le bateau de Thésée une fois toutes ses pièces remplacées. Toutefois, nous pensons que ce second cas ne pose pas plus de problèmes. Selon notre lecture, un être humain à 2 ans n'est pas le même faisceau que ce même être humain à 82 ans. S'il a conservé certains tropes qu'il avait à 2 ans, il y aura certaines similitudes entre les deux faisceaux : les tropes en question. Mais si l'ensemble des tropes du faisceau ont été remplacés, alors selon la théorie qui nous intéresse, la seule chose qu'auraient en commun l'individu de 2 ans et celui de 82 ans serait un lien temporel et causal : ils proviennent tous deux du même faisceau originel. La question de savoir si un être humain conserve certains de ses tropes tout au long de son existence entraînerait une discussion philosophique que nous n'avons évidemment pas la place d'avoir ici. Toutefois, il ne semble pas y avoir d'incohérence théorique dans cette conception de la persistance. La critique de Marmodoro et Mayr selon laquelle la théorie des tropes ne peut rendre compte de la persistance d'un objet malgré des changements au niveau des tropes le constituant est acceptée, mais notre réponse est : ce n'est pas grave.

D'autres versions de ces trois critiques existent, mais nous n'avons pas le loisir de toutes les présenter ici. La théorie des tropes semble donc bien s'en tirer face aux critiques métaphysiques qui lui sont

faites. Nous allons maintenant nous tourner vers sa compatibilité avec les connaissances scientifiques actuelles et nous demander si elle peut s'accompagner d'une position cohérente en philosophie des sciences.

## 3.2 Les tropes comme briques de construction : implications épistémologiques et métaphysiques

Avant de voir en quoi la théorie des tropes peut répondre au problème de l'exclusion causale, nous allons dire quelques mots de ses implications épistémologiques et métaphysiques afin de voir si elle est une position ontologique pertinente et légitime. Nous allons suggérer qu'elle apporte une contribution pertinente au débat épistémologique entre réalistes et antiréalistes, en offrant un appui théorique à l'antiréalisme scientifique dans son agnosticisme au sujet du statut ontologique des entités et processus inobservables postulées par les sciences fondamentales. En cela, elle est une position ontologique tout à fait légitime. Avant de préciser cette affirmation et de la défendre, nous allons brièvement présenter le débat qui oppose réalisme et antiréalisme scientifique<sup>52</sup>.

### 3.2.1 Réalisme et antiréalisme

Le débat entre réalistes et antiréalistes émerge d'une tension entre deux ensembles de considérations épistémologiques<sup>53</sup>. D'une part, certaines entités et processus qui apparaissent dans les théories scientifiques ne sont pas *observables* comme peuvent l'être des organes, des insectes ou des planètes, mais *postulés* par celles-ci pour expliquer les phénomènes qu'elles tentent de décrire et de prédire : c'est le cas des protéines, des gènes, des quarks ou encore des protons. Ces derniers peuvent être postulés pour plusieurs raisons : afin d'expliquer le comportement d'entités observables, ou encore pour prédire ou expliquer le comportement d'autres entités non observables déjà postulées auparavant. Certaines théories scientifiques, particulièrement en physique, sont donc peuplées de nombreuses entités postulées, mais jamais directement observées. Il arrive que le degré de croyance en leur existence augmente à la suite d'expériences<sup>54</sup>, mais la plupart restent inobservables. Cette différence de statut (observable et inobservable) entre les objets de plusieurs de ces théories pousse une partie de la communauté scientifique et

---

<sup>52</sup> Les termes « réalisme » et « antiréalisme » désignent des positions diverses dans toutes sortes de champs de recherche en philosophie et dans d'autres domaines, mais nous les utiliserons uniquement pour désigner le réalisme scientifique et l'antiréalisme scientifique tels que nous allons les définir dans la section qui suit.

<sup>53</sup> Il existe également un débat entre réalisme et antiréalisme métaphysique (un antiréalisme de cette espèce est parfois appelé *idéalisme*), qui discute l'existence d'un monde extérieur indépendant de l'esprit. Nous ne discutons pas celui-ci, et postulons un réalisme métaphysique, c'est-à-dire l'existence d'une réalité objective, indépendante de l'esprit : celle que les sciences étudient et dont elles tentent de prédire les phénomènes. Ce qui nous intéresse, c'est le statut des éléments que la science *postule* pour faire ses prédictions.

<sup>54</sup> On peut citer l'exemple du boson de Higgs, particule postulée dès les années 60 mais dont l'existence faisait débat et était très loin de faire l'unanimité parmi la communauté scientifique jusqu'en 2012, année où une équipe du Conseil européen pour la recherche nucléaire (CERN) a affirmé avoir des données suffisantes pour établir avec un haut taux de confiance l'existence d'une particule se comportant de la même façon que la particule postulée.

philosophique à adopter différents degrés de croyance en l'existence de ceux-ci. Autrement dit, il est question de leur accorder des statuts ontologiques différents.

Par ailleurs – et c'est là que la tension émerge – les succès remarquables de nos meilleures théories scientifiques poussent une autre partie de cette communauté scientifique et philosophique à les tenir pour au moins approximativement vraies, et donc à tenir les entités et les processus inobservables qu'elles postulent pour bien réels, ayant un statut ontologique équivalent à celui des entités macroscopiques observables. C'est cette posture que l'on nomme réalisme scientifique : une attitude épistémologique positive à propos du contenu de nos meilleures théories scientifiques, qui consiste à les considérer comme approximativement vraies et à croire en l'existence des éléments qui les peuplent. Bien entendu, il existe un spectre complet de positions réalistes. D'un côté de ce spectre on retrouve le simple réalisme structurel, position selon laquelle une certaine structure de la réalité est conservée au fil des théories successives décrivant la réalité. À l'autre extrémité, on trouve un réalisme naïf, qui consiste à tenir pour absolument vraies toutes nos meilleures théories scientifiques actuelles et à considérer que toutes les entités et processus qu'elles postulent existent réellement, et ont exactement le même statut ontologique que les entités macroscopiques observables. Si la première fait presque l'unanimité, la seconde n'est à notre connaissance défendue sérieusement par personne. Toutefois, on retrouve un grand nombre de positions intermédiaires tout au long du spectre.

Nous ne rentrerons pas dans le détail de chaque argument en faveur et à l'encontre du réalisme, car cela nous éloignerait trop de notre propos. Nous allons cependant en présenter brièvement deux, un pour et un contre, afin que le lectorat puisse cerner les motivations de chacun des camps plus clairement que si l'on se contentait d'en nommer les positions. Parmi les arguments classiques en faveur du réalisme, on peut citer le « No-miracle argument », formulé par Putnam en 1975. Il affirme que les succès prédictifs de la science ainsi que la remarquable précision de ceux-ci et des applications pratiques qu'elle produit (les technologies informatiques, ou encore l'ingénierie en général) ne peuvent être expliqués que par la vérité de ces théories, et l'existence des entités et processus qu'elles postulent, sans quoi il faudrait les attribuer à une forme de miracle. Le raisonnement est le suivant : si les théories sont fausses et que les entités qu'elles postulent n'existent pas, alors les succès prédictifs et les avancées technologiques qu'elles permettent sont dues à une chance inouïe ! Il semble plus raisonnable de les expliquer par le fait que les théories qui en sont responsables sont vraies – ou approximativement vraies.

Dans le camp des antiréalistes, un argument que l'on retrouve dans plusieurs textes a la forme d'une méta-induction pessimiste (on trouve ce terme chez plusieurs philosophes, et notamment chez Anjan Chakravartty, qui a beaucoup travaillé sur le débat entre réalistes et antiréalistes). Il affirme qu'au vue des réfutations successives de nombreuses théories scientifiques par le passé, il ne fait guère de sens de

s'attendre à mieux de celles avec lesquelles nous travaillons actuellement (Chakravartty 2017a). Cet argument est populaire notamment en raison des nombreux exemples qui semblent s'offrir à ses défenseurs : on ne compte plus le nombre de théories utilisées pendant un temps puis rejetée en raison de l'apparition d'une nouvelle théorie plus efficace (citons par exemple la théorie des humeurs, la théorie astronomique de Ptolémée, ou plus récemment la théorie du phlogistique). Pour comprendre les réserves que peuvent avoir les réalistes face à cet argument, il est utile de se pencher sur ce qu'est un raisonnement par induction, afin de voir s'il est possible d'en utiliser un dans le cadre du débat entre réalisme et antiréalisme.

Le terme « induction » désigne un type de raisonnement visant à extrapoler une conclusion tirée à propos d'un ou plusieurs éléments suite à une preuve ou à des observations à d'autres éléments pour lesquels une telle preuve ou de telles observations n'ont pas été faites. Cette extrapolation est justifiable de plusieurs façons. Dans le cas d'une preuve mathématique, par exemple, une preuve par induction requiert pour être valide ce que l'on nomme le « pas d'induction », qui se doit, lui, d'être démontré. Prenons comme exemple l'induction mathématique faible. Il s'agit d'un argument qui a la forme suivante :

$$[\text{Induction mathématique faible}] : [P_0 \ \& \ (\forall_n \mid P_n \supset P_{n+1})] \supset (\forall_n \mid P_n)$$

Où  $n$  désigne un entier naturel et  $P$  une propriété mathématique d'un entier naturel. On peut le lire ainsi : si le premier membre des entiers naturels a la propriété  $P$ , et si pour tout entier naturel, s'il a la propriété  $P$  alors l'entier naturel qui le suit a lui aussi la propriété  $P$  ; alors tous les entiers naturels ont la propriété  $P$ . Ce raisonnement est valide, et la conclusion est extrêmement forte : elle porte sur l'ensemble des entiers naturels. Toutefois, sa vérité dépend du cas de base d'une part, et du pas d'induction – la deuxième partie de la conjonction – d'autre part. C'est ce pas d'induction qui est généralement difficile à prouver. Sa démonstration dépend de la nature de  $P$ , et du fait que sa possession par un élément de l'ensemble implique sa possession par l'élément suivant.

Dans des domaines moins formels, il est difficile de trouver des utilisations correctes d'arguments par induction. Un exemple classique de mauvais raisonnement par induction est celui d'un individu qui, après avoir observé un grand nombre de chats noirs, formule une croyance du type « tous les chats sont noirs ». Ici, le pas d'induction est inexistant, ou à tout le moins, injustifié. Reprenons notre méta-induction pessimiste. Le(s) cas de base serai(en)t la(les) théorie(s) scientifique(s) réfutée(s), désormais en désuétude. Le pas d'induction, lui, est difficile à cibler avec précision : qu'est-ce qui peut justifier d'extrapoler les échecs des théories passées à nos théories actuelles et futures ? La qualité de l'argument est débattue et les partisans du réalisme scientifique affirment qu'il n'est pas cohérent. Toutefois, il reste un des arguments préférés des partisans de l'antiréalisme. Il existe bien d'autres arguments dans les deux camps, mais nous allons nous concentrer ici sur les conséquences que l'adoption d'une ontologie de tropes peut avoir sur le débat.



### 3.2.2 Réalisme et théorie des tropes

Est-ce qu'adopter une ontologie de tropes contraint à prendre position dans un tel débat ? Si oui, à quelle position doit-on adhérer ? Il n'existe pas une littérature très large sur la question. En revanche, on peut d'ores et déjà affirmer qu'il nous semble difficile de tenter d'associer une ontologie de tropes à un réalisme fort.<sup>55</sup> En effet, un tel réalisme n'aurait que faire de postuler des entités étranges comme les tropes, et ne s'en remettrait qu'aux théories scientifiques pour constituer son ontologie : celle-ci serait peuplée d'électrons, de quarks et de photons. Nous allons proposer qu'une ontologie de tropes s'arrime le mieux avec un antiréalisme. De plus, nous pensons qu'elle lui apporte l'appui métathéorique dont il a besoin pour justifier son agnosticisme au sujet des entités et processus inobservables postulés par les théories scientifiques. Mais avant, vérifions si elle peut aussi être vue comme compatible avec un réalisme faible.

Au premier coup d'œil, la théorie des tropes peut en effet sembler compatible avec une version très faible du réalisme – le réalisme structurel, évoqué brièvement dans la section précédente – que l'on peut formuler ainsi :

[Réalisme structurel]<sup>56</sup> : La nature des entités non observables est considérée de façon sceptique mais la structure de la réalité non observable, telle que représentée par des relations décrites dans nos meilleures théories, est réelle.

Face à des arguments comme la méta-induction pessimiste, cette position réaliste minimale tente de sauver les meubles : les théories scientifiques sont en effet réfutées les unes après les autres, et nous pouvons effectivement être instrumentalistes<sup>57</sup> quant aux entités postulées par celles-ci, mais une certaine *structure* de la réalité est toutefois conservée de fois en fois, et cela nous permet de rester des réalistes quant au discours scientifique. Les anciennes théories, selon cette perspective, sont vues comme des approximations des nouvelles, qui sont elles-mêmes des approximations de futures théories plus précises qui viendront un jour les remplacer. Cette position semble toutefois problématique à plusieurs égards. D'abord, le

---

<sup>55</sup> Voir à ce sujet le texte de Ilkka Niiniluoto, qui discute de la théorie des tropes dans le contexte d'autres débats en philosophie, et notamment dans le cadre du réalisme scientifique. Ses conclusions sont assez faibles et il se contente d'affirmer que la théorie des tropes n'implique pas un réalisme fort (Niiniluoto 2012, p. 448).

<sup>56</sup> Cette définition est une traduction libre de celle qu'on trouve dans l'article « Scientific Realism » de la *Stanford Encyclopedia of Philosophy* : « Structural realism is another view promoting selectivity, but in this case it is the natures of unobservable entities that are viewed skeptically, with realism reserved for the structure of the unobservable realm, as represented by certain relations described by our best theories » (Chakravartty 2017b)

<sup>57</sup> L'instrumentalisme est le nom parfois donné à l'antiréalisme qui s'oppose précisément à la dimension sémantique du réalisme scientifique, c'est-à-dire à l'idée selon laquelle les énoncés scientifiques à propos des processus et entités inobservables ont des valeurs de vérité – sont vrais ou faux. Les instrumentalistes voient au contraire ces énoncés comme des instruments de prédiction, qui n'ont d'utilité qu'au sein de la théorie. C'est d'ailleurs le terme qu'emploie Kim dans l'extrait cité au premier chapitre de ce mémoire (Kim 2010, p. 149).

concept de « structure » est assez vague. Quelle structure est effectivement conservée ? Un exemple utilisé par les défenseurs de cette position est celui de la loi de la gravitation d'Isaac Newton. Celle-ci a été révoquée par la théorie de la relativité générale d'Albert Einstein (la courbure de l'espace-temps y remplace la notion de force newtonienne), mais les réalistes structurels affirment que la première est une approximation de la seconde et est même assez efficace dans des conditions particulières : pour prédire le mouvement d'objets sur Terre, par exemple. Toutefois, le philosophe Thomas Kuhn, entre autres, considère que les travaux d'Einstein ne constituent pas simplement un ajustement par rapport à ceux de Newton, mais une réfutation complète de ceux-ci, tout comme les constats de Copernic et de Galilée rejetaient complètement la théorie des sphères de cristal de Ptolémée, malgré les succès prédictifs que celle-ci rencontrait jusqu'alors (Kuhn 1996).

Les réalistes structurels semblent de plus confondre les conséquences métaphysiques et pragmatiques de ces réfutations, de ces changements de paradigme. Sur le plan métaphysique, la réfutation est totale : le système de Ptolémée ne décrit pas l'univers, il n'existe pas de sphères de cristal. Les forces newtoniennes n'existent pas elles non plus. Sur le plan pragmatique, toutefois, les conséquences sont effectivement minimales sur les applications pratiques des théories : les ingénieurs continuent d'utiliser les forces newtoniennes et semblent très bien s'en sortir. En fait, c'est précisément ces deux éléments que les réalistes confondent : le fait qu'une nouvelle théorie soit l'approximation de la précédente d'une part, et le fait que certains résultats de l'ancienne théorie soient des approximations suffisantes de ceux de la nouvelle (au point de continuer à pouvoir être utilisés dans des domaines pratiques et des champs de recherche ne portant pas directement sur des éléments fondamentaux<sup>58</sup>) d'autre part. Le premier constat est évidemment faux. Comme le note Kuhn, ce sont les conceptions même d'espace et de temps qui ont été révisées, Einstein leur a fait perdre leur caractère absolu, jusque-là pris pour acquis (Kuhn 1996). Le second constat, quant à lui, est évidemment vrai. Cela ne pose pas de problème particulier : il suffit d'accepter que nos meilleures théories scientifiques ne sont pas des descriptions de la réalité, mais des outils prédictifs. Ceux-ci sont donc voués à s'améliorer constamment et à modifier leurs paradigmes pour suivre l'évolution des recherches fondamentales.

Il est peu probable que la théorie des tropes puisse fournir au réalisme structurel le fondement ontologique qui l'aiderait à résoudre les difficultés auxquelles il fait face. Il est donc plus fécond d'associer

---

<sup>58</sup> En fait, la majorité de l'ingénierie se contente du système et des équations newtoniennes. Comme contre-exemple, outre les recherches en physique et en astrophysique qui portent explicitement sur la théorie, on peut citer les systèmes d'évaluation temporelle ou de localisation GPS des satellites qui doivent procéder à des ajustements prenant en compte la relativité générale afin de garantir l'exactitude de leurs résultats. Ceux-ci ne peuvent donc pas se contenter des paramètres physiques newtoniens, car ils les rendraient inutiles.

celle-ci à l'antiréalisme. Les antiréalistes affirment qu'il n'est pas nécessaire pour maintenir un réalisme métaphysique (l'existence d'une réalité indépendante de l'esprit) de croire à l'existence des entités ou processus non observables postulés par les théories scientifiques ; il suffit de rester agnostiques à leur sujet. Or, il semble que les tropes, en tant que particuliers abstraits, permettent aux antiréalistes d'échapper à un engagement ontologique envers les entités et processus inobservables postulés par les théories scientifiques, tout en gardant une ontologie cohérente, et ce même lors de changements de paradigme dans un domaine de recherche. En philosophie de l'esprit, par exemple, concevoir les états mentaux comme des tropes permet d'accommoder les changements de paradigme que seront amenées à traverser les neurosciences sans devoir revoir constamment notre ontologie.

Une objection peut venir en tête lorsqu'on considère la position du théoricien des tropes antiréaliste. N'est-ce pas là une contradiction dans les termes ? Affirmer que nos meilleures théories scientifiques sont seulement des outils prédictifs ne revient-il pas à abandonner toute forme de projet ontologique ? Dans ce cas, ne devons-nous pas également abandonner la théorie des tropes ? En fait, comme le notent Heil et Martin dans leur article « The Ontological Turn », toute forme d'antiréalisme n'est pas incohérente, mais il est erroné de croire qu'une position antiréaliste dispense son défenseur d'une justification ontologique (Martin et Heil 1999, p. 35). Pour eux, un antiréaliste doit simplement préciser sa position en mentionnant la portée de celle-ci (sur quel domaine porte son antiréalisme), ainsi qu'en définissant son ontologie. La proposition des auteurs est que le rôle de l'ontologie est de s'inspirer de l'état actuel des connaissances empiriques et théoriques – c'est-à-dire à la fois des meilleures théories et résultats scientifiques disponibles, et des débats ayant lieu en métaphysique et en épistémologie contemporaines – pour imposer des contraintes et des normes aux sciences : « ontology's abstractness provides what amount to constraints or placeholders for certain kinds of less abstract model in science » (p. 54).

Pour Heil et Martin, la théorie des tropes peut fournir des lignes directrices qui peuvent guider les recherches empiriques. Cela donne donc un bon aperçu de l'idée qu'ils se font de la relation entre sciences et ontologie. En particulier, il semble clair qu'il ne faut pas négliger le rôle de cette dernière, mais bien puiser en elle des limites et des directions théoriques qui peuvent aider à aiguiller la recherche. En cela, la théorie des tropes offre une solution propice à la conciliation d'un antiréalisme scientifique et d'un positionnement ontologique.

On voit donc que la théorie des tropes offre un support ontologique aux agnostiques de l'inobservable, et qu'être antiréaliste ne condamne pas nécessairement à l'agnosticisme ontologique. S'il ne s'agit pas d'un argument décisif en faveur de la théorie des tropes, cela contribue à la rendre attrayante.

Revenons-en à présent au problème de l'exclusion, avec lequel nous avons commencé, et voyons dans le détail en quoi la théorie des tropes y apporte une solution réductionniste satisfaisante.

### 3.3 La théorie des tropes, une solution réductionniste

Les différentes théories des tropes ont ceci en commun qu'elles pointent toutes en direction d'une solution réductionniste au problème de l'exclusion. Dans cette section, nous allons nous pencher sur ce type de solution, et sur les raisons qui nous poussent à classer ces théories parmi la famille des solutions réductionnistes.

Si l'on adopte la perspective ontologique d'une théorie des tropes, il nous faut reconsidérer les thèses sur lesquelles nous nous sommes penchés au premier chapitre. Dans certaines d'entre elles, le terme « propriété » est employé pour référer aux particuliers abstraits – les tropes – et dans d'autres il s'agit de classes de ressemblance de tropes. C'est en révélant cette ambiguïté que nous allons résoudre la contradiction à laquelle nous faisons face. En suivant l'ensemble de nos axiomes de départ, nous arrivons au raisonnement suivant :

- (1) M cause M\*. (Par [Efficacité causale du mental])
- (2) M\* survient sur P\*, sa base de survenance physique. (par [Survenance du mental])
- (3) M cause M\* *en causant* sa base de survenance P\*.
- (4) M est une cause de P\*. (par 3)
- (5) P\* a une cause physique P. (par [Complétude du physique] et 4)
- (6)  $M \neq P$ .
- (7) P\* a deux causes. (par 4, 5 et 6)
- (8) Il faut éliminer M ou P comme cause de P\*. (par [Exclusion], si l'on n'est pas dans un cas authentique de surdétermination)
- (9) M est abandonnée comme cause de P\*. (par [Complétude du physique] et 8)

Au-delà du fait que (9) contredit (4), et que nos thèses sont donc inconsistantes, rappelons que nous ne voulons pas de (9). Nous voulons que les propriétés mentales *soient* efficaces causalement. Or, voir les propriétés comme des tropes nous permet de conserver cette efficacité causale. Il suffit pour cela de clarifier

la notion de propriété telle qu'employée dans les thèses. Reprenons-les une à une afin de voir comment elles s'arriment à la position ontologique développée dans ce chapitre.

[Efficacité causale du mental] : Certains événements en causent d'autres du fait qu'ils sont l'instanciation d'une propriété mentale.

Ici, les deux notions que recouvre le terme « propriété » sont entremêlées. Ce qui doit être compris, si l'on accepte une ontologie de tropes, c'est que certains tropes qui sont causalement efficaces tombent *incidemment* dans une classe de ressemblance que l'on nomme « mentale ». Toutefois, ce n'est pas en vertu de son appartenance à une telle classe que le trope en question est efficace causalement. Un trope est efficace en tant que particulier abstrait. C'est ce que signifie ici le terme « propriété », il ne renvoie pas à une classe de ressemblance. Comme le note Maurin : « Tropes are not causally relevant in virtue of belonging to this or that similarity class, because they are not mental (or physical) in virtue of belonging to some such class. A trope is what it is, primitively. It is in virtue of being what it is that it is causally relevant (if it is) » (Maurin 2008, p.39). Si l'on considère les propriétés comme des tropes, alors il n'est rigoureux de parler de propriété *mentale* ou *physique* que dans la mesure où le trope que l'on considère appartient à une classe de ressemblance de tropes que nous nommons « tropes mentaux » ou « tropes physiques ». L'efficacité causale du mental demeure donc, mais dans un sens trivial, déjà capturé par le paradigme physicaliste et par la fermeture causale du physique : ce dont nous parlons lorsque nous décrivons nos états mentaux, l'élément du monde désigné à l'aide du lexique du mental, est également descriptible à l'aide de termes physiques : comme un ensemble de processus biologiques, chimiques ou électriques, dans notre cas. Il n'est donc pas pertinent de se demander : « en vertu de son appartenance à quelle classe de ressemblance ce trope est-il efficace causalement ? ».

[Survenance du mental] : Les propriétés mentales surviennent sur des propriétés physiques.

Si l'on considère les propriétés comme des tropes, il semble peu vraisemblable de considérer que cette thèse porte sur ces derniers. En effet, on l'a dit, tout trope mental *est* un trope physique ; affirmer qu'un trope survient sur lui-même est incohérent, car la relation de survenance est une relation de dépendance asymétrique. La propriété survenante dépend de la propriété subvenante, mais non l'inverse, c'est donc une relation qu'aucune entité ne peut entretenir avec elle-même. Il ne serait pas plus cohérent d'affirmer qu'un trope doit survenir sur un autre trope puisque les théoriciens des tropes refusent les niveaux de réalité. Le but de [Survenance], dans notre problème d'origine, est de réconcilier [Efficacité causale du mental] et [Irréductibilité], en faisant en sorte qu'une propriété puisse être pertinente causalement sans être physique. Or,

comme le dit Robb<sup>59</sup>, son monisme de trope est plus éclairant car il explique également *comment* ladite propriété est efficace (ce que ne faisait pas ce qu'il appelle « la solution survenante ») : elle est elle-même physique. Il écrit : « by showing that mental causation is really just a kind of physical causation, the trope solution provides the “how” of mind-body interaction » (Robb 1997, p.189). Alors faut-il abandonner [Survenance], ou bien considérer qu'elle est impliquée par le monisme de trope ? Pour Robb, il semble que l'idée de survenance s'applique davantage aux classes de ressemblance de tropes qu'aux tropes eux-mêmes. Pour lui, la classe de ressemblance mentale à laquelle appartient un trope ne se réduit pas à la classe de ressemblance physique à laquelle appartient ce même trope. Toutefois, la classe physique est un sous-ensemble de la classe mentale : « the mental type is not itself a physical type [...] though of course it has many physical types as subsets : these are just the physical types that “realize” the mental one » (Robb 1997, p. 188). Rappelons que Robb choisit le terme « type » pour rester neutre sur la question de l'existence des universaux, alors que nous nous servons du terme « classe ». Ici, Robb emploie le terme « réalise » entre guillemets car il ne considère pas que les classes – qui sont des catégorisations a posteriori – surviennent les unes sur les autres comme les propriétés surviennent les unes sur les autres dans les théories de la survenance. Toutefois, cela aide à se représenter le type de relation qu'entretiennent les tropes avec les classes et les classes entre elles. Par ailleurs, puisque nous avons adopté [Identité] et que les propriétés mentales sont identifiées avec des propriétés physiques, on pourrait croire, à première vue, que [Survenance du mental] est en quelque sorte garantie par [Identité]. Cette conclusion est cependant fautive car, comme nous venons de le mentionner, cela voudrait dire qu'il est correct de croire qu'une propriété survient sur elle-même, ce qui est ontologiquement invraisemblable.

[Irréductibilité] : Les propriétés mentales ne sont ni réductibles ni identiques à des propriétés physiques.

Dans cette thèse, le terme « propriété » est encore une fois ambigu. Robb propose un *monisme* de trope, et tout trope mental est donc identique à un trope physique. Ainsi, on ne peut garder cette thèse que s'il y est question de *classes de ressemblance*. D'ailleurs, il faut la conserver : on ne veut en effet pas tomber dans l'identité de types, au risque de voir l'argument de la réalisabilité multiple revenir à la charge. L'idée derrière cette thèse, selon notre perspective, est donc qu'une classe de ressemblance de tropes mentaux (par exemple la douleur) n'est pas réductible ou identifiable à une classe de ressemblance de tropes physiques (la stimulation des fibres C). Plutôt, au sein de la classe de ressemblance mentale de la douleur, on retrouve des tropes qui appartiennent à une multitude de classes de ressemblance physiques différentes. Robb écrit : « a

---

<sup>59</sup> Cela est aussi compatible avec la position de Heil, toutefois Robb est le celle à employer l'expression « monisme de trope » à notre connaissance – sans doute car Heil préfère le terme « mode ».

given mental type is a set of physical tropes. But multiple realizability entails that these physical tropes do not themselves resemble one another in the way that members of a *physical* type must: they will be wildly dissimilar physically » (Robb 1997, p. 188).

[Exclusion] : Un évènement ne peut généralement pas avoir plus d'une cause suffisante.

Pour discuter cette thèse, il faut se demander ce qui joue un rôle dans les relations de causalité. Nous l'avons dit au premier chapitre, nous considérons que ce sont les évènements qui se causent entre eux. Si nous clarifions cette affirmation en prenant en compte notre ontologie de tropes, nous considérons que ce sont les tropes, des propriétés particulières, et non les types de propriétés ou les classes de ressemblance, qui jouent un rôle causal. Comme le note Robb : « they are the "properties" that are causally relevant in causal relations; they are, as I shall say for short, the properties of causation » (Robb 1997, p. 187).

[Complétude du physique] : Si un évènement a une cause à  $t$ , il a une cause physique à  $t$ .

Ici, la notion de propriété n'est pas mentionnée directement, mais on pourrait reformuler la thèse ainsi : pour tout évènement, si un trope joue un rôle causal dans sa production, alors un trope appartenant à la classe des tropes physiques doit jouer un rôle causal dans sa production. Cela est consistant avec la remarque de Robb selon laquelle les tropes sont les propriétés responsables des relations causales. Pour Robb, puisque les tropes mentaux sont tous des tropes physiques (selon son monisme de trope), alors [Efficacité causale du mentale] est consistante avec [Complétude du physique] : la première postule que certains tropes mentaux sont efficaces causalement (souvenons-nous : elle n'affirme pas qu'ils sont efficaces *en vertu du fait* qu'ils sont des tropes mentaux, elle affirme simplement qu'il existe des tropes mentaux efficaces causalement), et la seconde affirme que tout trope efficace causalement est un trope physique. Grâce au monisme de trope, nul besoin de chercher comment agencer ces deux thèses en postulant une relation de dépendance comme la survenance : tout trope mental est un trope physique.

La théorie des tropes propose donc de réconcilier les différentes thèses du problème de l'exclusion en clarifiant les usages du terme « propriété » : les propriétés efficaces causalement sont des tropes, et elles sont efficaces causalement en vertu de ce qu'elles *sont* fondamentalement. Certains tropes appartenant à des classes de ressemblance mentales sont efficaces causalement, mais puisqu'ils appartiennent également à des classes de ressemblance physiques, il n'y a pas d'explication supplémentaire à fournir sur la façon dont ils sont efficaces causalement : la causalité mentale n'est qu'une forme de causalité physique. Voyons à présent en quoi la théorie des tropes est une position réductionniste pertinente.

### 3.4 La réalisabilité multiple : bête noire du réductionnisme?

En cela qu'elle rejette un dualisme ontologique et ne conserve une distinction entre mental et physique qu'au sein du domaine descriptif, la théorie des tropes peut bien être qualifiée de solution réductionniste au problème de l'efficacité causale du mental. Pour évaluer sa pertinence, on peut donc se demander comment elle se débrouille face aux critiques adressées à cette famille de positions. Nous allons considérer une des critiques les plus abordées dans la littérature sur le réductionnisme : l'argument de la réalisabilité multiple.

La possibilité qu'un même type d'état mental puisse être réalisé par plusieurs processus différents (biologiques, chimiques, ou autre) a d'abord été soulevée comme une critique des positions de l'identité des types (et du projet réductionniste plus généralement) dans les années 60 par Putnam. Il est intéressant de considérer en détail cet argument car la théorie des tropes lui apporte une réponse intéressante et potentiellement salvatrice pour la position réductionniste. Nous allons d'abord nous intéresser à la structure de l'argument et aux différentes prémisses, avant de considérer la réponse qu'offre la théorie des tropes.

#### 3.4.1 L'argument de la réalisabilité multiple

L'argument, dans sa forme classique, prend la forme d'un *modus ponens*. Il repose sur une prémisse principale, la thèse de la réalisabilité multiple. Il s'agit de la thèse selon laquelle un même type mental peut être réalisé par différents types physiques. Nous allons la formuler de façon plus générale ici, en utilisant un quantificateur *existential*.<sup>60</sup>

[Réalisabilité Multiple] : Au moins certains types mentaux sont réalisables par différents types physiques.

À cette prémisse principale, que nous allons discuter par la suite, on ajoute l'implication matérielle suivante :

[Prémisse mineure] : Si un type mental est réalisable par plusieurs types physiques, alors il ne peut pas être identifié à un de ces types en particulier.

À partir de ces deux prémisses, nous sommes amenés à une version (quantifiée ici encore avec un quantificateur existentiel) du conséquent de cette implication matérielle :

[Conclusion] : Au moins certains types mentaux ne sont pas identifiables à des types physiques.

---

<sup>60</sup> En effet, dans sa forme classique, l'argument doit mettre en péril les théories de l'identité des types, et il est suffisant qu'il existe au moins un type d'états mentaux qui ne soit pas identifiable à un type d'états physiques pour que celle-ci soit falsifiée.



Cette conclusion est suffisante pour rejeter les théories de l'identité des types, qui ont une forme universelle (et sont donc falsifiées par l'existence d'un seul contre-exemple). La deuxième prémisse ne semble pas problématique, pas plus que la validité de l'argument. Il semble donc bien que l'efficacité de celui-ci repose sur la vérité de la première prémisse. Prenons le temps d'examiner chacune de ces deux prémisses.

Pour ce qui est de la prémisse mineure, il faut préciser la relation d'identité qui est en jeu dans celle-ci. La réduction d'un domaine à un autre – selon certains détracteurs de la position réductionniste à tout le moins<sup>61</sup> – implique non seulement une identité d'occurrences telle que nous l'avons définie plus haut (cf section 2.2.2) mais également une identité de types. Une telle identité correspondrait à une réduction de ce que Jerry Fodor appelle les « special sciences » aux sciences physiques (Fodor, 1974). Cette réduction consisterait à être capable d'exprimer tout phénomène, événement, comportement d'entités ou d'objets à l'aide de prédicats et de termes appartenant à ces sciences physiques (possiblement à l'aide de lois ponts, reliant les deux domaines).

Si la thèse de la réalisabilité multiple est fautive, la relation d'identité n'est pas en danger : on peut identifier chaque type mental au type physique qui le réalise (et l'identification n'est qu'une question empirique, qui est liée au progrès du domaine en question – les neurosciences dans le cas du mental). Toutefois, si la thèse est vraie, et qu'un même type mental peut être réalisé par différents types physiques, alors il faudrait pour que la prémisse mineure soit fautive, et donc que le mental soit identique au physique, que la base physique soit une disjonction de tous les types possibles pouvant réaliser le type mental. Cela semble problématique, et des philosophes comme Fodor affirment qu'il est hautement improbable que des théories physiques puissent exprimer de façon exhaustive la liste de tels types de façon à garantir l'identité de la disjonction avec le type mental. En effet, il semble que le grand nombre de réalisations possibles rendent la formulation d'une telle disjonction au mieux trop compliquée et au pire impossible.

Il semble donc que l'argument repose sur la prémisse principale, la thèse de la réalisabilité multiple. Celle-ci est parfois considérée comme étant une question empirique au sujet de laquelle il est difficile de se prononcer. Block et Fodor affirmaient d'ailleurs dans leur article de 1972 : « Of course, it is possible that the type-to-type correspondences required by behaviorism or by physicalism should turn out to obtain. The present point is that even if behavioral or physical states are in one-to-one correspondence with psychological states, we have no current evidence that this is so; hence we have no warrant for adopting philosophical theories which require that it be so » (Block et Fodor 1972, p. 162). Nous n'argumenterons pas en faveur ou à

---

<sup>61</sup> C'est le cas de Jerry Fodor qui affirme dans son article de 1974 : « reductivism is the conjunction of token physicalism with the assumption that there are natural kind predicates in an ideally completed physics which correspond to each natural kind predicate in any ideally completed special science » (p. 100).

l'encontre de cette thèse, mais nous considérerons dans les prochaines sections les conséquences qu'entraînerait sa vérité ou sa fausseté. Pour notre propos, il est suffisant de mentionner que cette thèse ne fait pas l'unanimité et nous la traiterons donc en tant qu'hypothèse. Parmi les raisons qui poussent les philosophes à douter de sa vérité, on peut mentionner une observation faite par Kim dans son article « Multiple Realization and the Metaphysics of Reduction » quant au succès des recherches en neurosciences qui comparent les états neuronaux d'espèces différentes : si les états mentaux étaient effectivement réalisables par différents processus physiologiques chez différentes espèces, il serait étrange que des conclusions puissent être tirées au sujet d'une espèce à partir des résultats obtenus grâce à l'étude d'une autre. Il ne serait donc plus possible de concevoir la psychologie comme une science dont l'objet d'étude est unifié (Kim 1992). Nous n'allons pas examiner en détail la plausibilité de la thèse de la réalisabilité multiple car cela nous éloignerait quelque peu de notre propos. Toutefois, elle est encore aujourd'hui abondamment discutée et il est donc important de considérer ses impacts sur le réductionnisme. Voyons maintenant comment une ontologie de tropes peut répondre à cette objection contre la réduction du mental au physique.

### **3.4.2 Quelle réponse pour le réductionnisme ?**

Si les états mentaux particuliers des individus sont des tropes, alors l'écueil de la réalisabilité multiple n'en est plus un, car les différents tropes d'un même type d'état mental (c'est-à-dire la douleur d'un humain, la douleur d'un chat, celle d'une espèce extraterrestre, etc.) appartiendront à différentes classes de ressemblance de tropes physiques (celle des stimulations de fibre C chez les humains et les chats, des gonflements de certaines cavités sous les pieds chez une espèce extra-terrestre, etc.) tout en appartenant à une seule et même classe de ressemblance de tropes mentaux, la classe de tous les particuliers qui sont des douleurs. La thèse de la réalisabilité multiple – si elle est vraie – est donc respectée. Bien que cette réponse puisse sembler simple, elle nécessite quelques précisions pour ne pas être confondue avec d'autres solutions au problème posé par l'argument de la réalisabilité multiple. La réponse qu'une ontologie de tropes propose *n'est pas* une identité de types, comme nous l'avons précisé à la section 2.2.2. Il ne s'agit pas non plus d'une tentative de réfutation de la thèse de la réalisabilité multiple. Or, c'est précisément cela que ses détracteurs (en tous cas ceux qui se servent de l'argument de la réalisabilité multiple à son encontre) reprochent à la position réductionniste : l'ajout d'une identité de types à l'identité d'occurrences. Toutefois, les défenseurs d'une ontologie de tropes n'auront aucun problème à rejeter l'identité des types et à admettre la possibilité d'une multitude de tropes différents dans leurs structures physico-chimiques appartenant cependant à la même classe de tropes mentaux. Ils affirment simplement qu'un tel dualisme n'est qu'un dualisme théorique, conceptuel, et non un dualisme ontologique. Il n'implique en aucun cas ce qu'ils appellent des « ontological additions to our world » (Heil et Robb 2003, p.188). Les types – ou comme ils les appellent, les classes – ne sont que des manières abstraites de parler des entités qui peuplent le monde (les tropes).

Nous l'avons dit plus tôt, l'argument de la réalisabilité multiple est également mis de l'avant pour défendre une position fonctionnaliste, et non seulement pour critiquer le réductionnisme. La raison en est que puisqu'il semble impossible (si la thèse de la réalisabilité multiple est vraie) d'identifier un type mental à un type physique, la théorie de l'identité des types s'avère insuffisante pour donner une explication satisfaisante des phénomènes mentaux. En effet, réduire ces derniers à leurs manifestations physiologiques individuelles manquerait de généralité (car cela empêcherait de pouvoir parler de mêmes états d'une espèce ou d'un individu à l'autre), et il est exclu *ex hypothesi* de les réduire à un seul type ou à une disjonction de types physiologiques. Toutefois, Block et Fodor jugent que le fonctionnalisme dispose des ressources suffisantes pour exprimer cette généralité des états mentaux. Or, la solution offerte par la théorie des tropes est d'accepter les conclusions de l'argument de la réalisabilité multiple en désambiguïsant ses prémisses afin d'éviter le retour de l'écueil du dualisme ontologique. Est-ce à dire que cela implique que la théorie des tropes doit entériner la position fonctionnaliste ? En quelque sorte oui. Comme nous l'avons montré à la fin du second chapitre, il est d'ailleurs possible d'articuler une conception fonctionnaliste de la notion de propriété (conçue comme classe de ressemblance) avec une ontologie de tropes.

La théorie des tropes semble donc être une avenue particulièrement intéressante pour le réductionnisme. Avant de conclure ce mémoire, nous allons nous tourner vers une critique portant spécifiquement sur cette utilisation de la théorie pour résoudre le problème de l'exclusion.

### **3.5 Le retour du problème de l'exclusion**

Dans cette section, nous allons nous intéresser à une critique adressée à la théorie défendue par Robb et Heil. Pour la résumer, cette critique affirme que le problème de l'exclusion resurgit dès lors qu'on a désambiguïté la notion de propriété, et de façon plus précise que jamais : si les tropes sont responsables de relations causales générant nos actions, est-ce en vertu de leur appartenance à une classe de tropes physiques ou à une classe de tropes mentaux ? Si c'est en vertu de la seconde, il semble qu'il faille justifier le rejet de [Complétude du physique], et si c'est en vertu de la première, alors il semble que l'appartenance du trope à la classe mentale soit superflue, que son importance soit exclue par le rôle joué par ses caractéristiques physiques. Cette critique est développée par Paul Noordhof dans un article de 1998. Nous allons la présenter en détail puis voir la réponse proposée par Robb.

Le problème de l'exclusion, rappelons-le, montre qu'en postulant à la fois l'unicité de cause (nous ne voulons pas de surdétermination causale systématique) et la complétude du domaine physique, le rôle causal joué par une propriété mentale qui surviendrait sur une propriété physique ou serait réalisée par une telle propriété est exclu par celui joué par cette dernière. La propriété mentale risque alors d'être épiphénoménale, au mieux. La théorie des tropes propose de regarder les choses autrement : il n'y a pas deux propriétés (au

sens particulier du terme) impliquées dans une relation causale entre états mentaux ou entre un état mental et une action, mais bien un seul trope. C'est ce trope qui appartient à (au moins) deux classes de ressemblance distinctes : l'une physique et l'autre mentale. L'appartenance à une classe de ressemblance, nous l'avons vu à la fin du second chapitre, est déterminée par le fait que le rôle fonctionnel décrit dans la définition de ladite classe est rempli par ce trope. Rappelons enfin que bien qu'elle soit dualiste en ce qui concerne les types (les classes de ressemblance), le fait que la théorie des tropes ne postule qu'un seul type d'entités en fait une position moniste.

Or, une question épineuse peut se poser pour certaines positions monistes lorsque vient le temps d'expliquer les mécanismes des relations causales auxquelles les entités qu'elles décrivent prennent part. C'est notamment le cas du monisme anomal de Davidson, qui propose un monisme des événements pour régler la question de l'efficacité causale du mental. Puisque sa théorie consiste à affirmer que tous les événements mentaux sont des événements physiques, la question a la forme suivante : l'évènement est-il pertinent causalement en vertu du fait qu'il est un événement physique ou mental? Nous ne rentrerons pas dans le détail de la position de Davidson ni dans celui des défenses qui peuvent en être fait. Plutôt, nous allons nous demander si la même question peut être posée dans le cas des tropes. Robb propose de faire des tropes les propriétés de la causalité, « the properties of causation » (Robb 1997, p. 187). De plus, il affirme afin de respecter [Complétude du physique] que tout trope mental est un trope physique. Soit. Mais la question qui est posée à Davidson peut-elle être posée à Robb également ? En vertu de quoi le trope est-il efficace causalement? En vertu du fait qu'il est un trope mental ou physique ? C'est là le cœur de la critique de Noordhof. Celui-ci affirme que le problème de l'exclusion resurgit au niveau des classes de ressemblance. Or, pour lui, une solution efficace à un problème philosophique ne devrait en aucun cas faire advenir un problème similaire qu'il faudrait régler à son tour. Une telle solution ne serait pas efficace et ne serait donc pas une bonne candidate. C'est ce qu'il appelle la contrainte de la bosse dans le tapis (« bulge in the carpet constraint », p. 222). Cette contrainte semble raisonnable. L'enjeu est donc de savoir si le problème de l'exclusion refait effectivement surface lorsque l'on adopte une ontologie de tropes pour tenter de le résoudre. Ceci nous ramène donc à la question de savoir s'il est cohérent de poser la question : « en vertu de quoi le trope joue-t-il un rôle causal? ».

Noordhof attaque la tentative de Robb pour esquiver la question du retour du problème de l'exclusion à partir d'une réponse anticipée de ce dernier dans son article de 1997. En effet, Robb pense que toutes les théories des propriétés échappent à la critique, par opposition à des théories sur les particuliers complexes (comme la théorie de Davidson). Il écrit : « there is a danger of vicious regress if we allow the sort of question that the objector raises [...] I see no reasons for allowing the problem to transfer to properties, whether or not we view properties as tropes » (Robb 1997, p. 191). La réponse de Noordhof est de contester cette idée en

considérant deux conceptions de la notion de propriété admettant des parties distinctes : les propriétés complexes et les propriétés en tant qu'instanciations.

My house burns down. It is quite legitimate to ask which aspect of air was responsible. The answer is that the air was causally relevant in virtue of being part oxygen. So it seems that complex properties do have aspects concerning which one can ask "Was that responsible?", namely, their constituents. In Robb's own example a red ball is dropped on a sheet of metal, causing a dent. He suggests that, once we have recognized that there was a dent because the ball had mass, we should not ask the further question "But was the mass causally relevant qua mass?". This appears wrong. Suppose that cricket balls would have dented the metal whereas tennis balls (with less mass) would not. We might allow that the red ball causes a dent qua having mass, but go on to say that, more specifically, it was qua having a mass of 5lb. rather than just qua mass. It also seems that, if two properties stand as determinate to determinable, we can ask "Which is relevant?". So we have two perfectly clear senses in which properties can have aspects which raise questions of causal relevance. (Noordhof 1998, p. 223)

Ces exemples semblent pertinents et pour nous, cet aspect de la critique de Noordhof fait mouche. Pour le comprendre, il suffit de considérer l'exemple de l'air. Dans un tel cas, nos connaissances en chimie et en physique à elles seules suffisent à nous en apprendre assez sur ce gaz pour en isoler les composants et leurs caractéristiques, ce qui nous renseigne sur quels aspects de l'air sont responsables de tel ou tel de ses comportements. Il est donc non seulement pertinent de demander « en vertu de quoi l'air est-il responsable de la combustion de ma maison ? », mais il est également assez facile d'y répondre de manière simple. Il semble alors que pour une telle conception de la notion de propriété – dans laquelle être de l'air, être de l'oxygène, etc., sont considérés comme des propriétés – le problème de l'exclusion menace en effet de refaire surface à chaque fois que l'on isole un aspect de la propriété dont on considère le rôle causal. De plus, s'il est légitime de demander quel aspect de l'air est pertinent pour expliquer son rôle causal, il semble cohérent de poser la même question au sujet de l'oxygène et de continuer notre enquête. Or, cette façon de se pencher sur le problème fait écho à une conception que les théoriciens des tropes souhaitent laisser de côté : la vision en strates de la réalité, dont nous avons parlé dans la section 2.2.1. Il n'est donc pas possible de rejeter en bloc l'objection du retour du problème de l'exclusion pour *toute* conception de la notion de propriété, contrairement à ce que Robb (1997) affirme. Pour empêcher le problème de resurgir, n'importe quelle conception des propriétés ne fait pas l'affaire. Il nous faut expliquer en quoi la conception des propriétés comme particuliers abstraits que met de l'avant la théorie des tropes est en mesure de contrer la version du problème de l'exclusion causale formulée par Noordhof, et en quoi cela la distingue de ses concurrentes.

C'est d'ailleurs plutôt cette stratégie que choisira Robb par la suite. Dans un article ultérieur, paru en 2013, il abandonne l'idée selon laquelle la difficulté rencontrée par la théorie de Davidson ne s'applique à aucune conception de la notion de propriété. À la place, il affirme que poser une telle question démontre une mauvaise compréhension de ce qu'est un trope. Pour lui, le trope n'a pas de parties constituantes, n'a pas lui-

même de propriétés, et ne possède pas tel ou tel pouvoir causal. Plutôt, il *constitue* la réalité, il est lui-même une propriété (au sens particulier du terme), il est lui-même tel ou tel pouvoir causal (rappelons-nous que pour Robb et Heil, un trope est à la fois une propriété dispositionnelle et catégorique; cf. section 2.1.3). Si l'on reformule la question en précisant la notion de trope, on commence à voir en quoi la critique de Noordhof ne rend pas justice à la théorie des tropes. Il ne semble pas cohérent de demander : quelle partie de ce pouvoir causal est responsable de la relation causale ? Noordhof semble confondre les tropes d'une part, et les objets ou entités macroscopiques postulés, nommés et discutés par les êtres humains, d'autre part (les molécules d'air, les atomes d'oxygène, etc.). Pour le théoricien des tropes, cette deuxième catégorie regroupe en fait des amas de tropes isolés arbitrairement du reste de la réalité par les observateurs que nous sommes, caractérisés et nommés par nos soins. En revanche, il est cohérent de demander : quel trope de cet objet est responsable de la relation causale dans laquelle il est impliqué ? En isolant le bon trope, on désignerait de fait le pouvoir causal en question. Pour Robb, donc, la question « en vertu de quoi... ? » ne se (re)pose pas. Si l'on associe la théorie des tropes à l'antiréalisme scientifique dont nous avons parlé un peu plus tôt, il ne serait pas question d'enlever au discours scientifique sur les gaz et leurs caractéristiques sa pertinence explicative ; mais d'un point de vue ontologique nous regarderions la situation autrement. Il y aurait un trope responsable du début de l'incendie, et un trope responsable de la combustion pour chaque atome d'oxygène prenant feu. Ces tropes sont indivisibles. Comme Robb le mentionne dans sa réponse de 2013 : « there are no tropes of tropes » (p. 221). Ainsi, il n'est vraiment pas cohérent avec la manière dont la réalité est dépeinte par la théorie des tropes de poser la question : « en vertu de quoi ce trope est-il pertinent causalement ? ».

Noordhof affirme dans son texte qu'il ne voit pas bien en quoi la régression dénoncée par Robb dans son article de 1997 serait vicieuse. Pour lui : « either there are aspectless properties or there are not » (Noordhof 1998, p. 221). S'il n'y en a pas, dit-il, la régression n'est pas vicieuse et l'on peut – on doit – reposer la question à chaque niveau. Or, c'est précisément ce qui cloche avec la critique de Noordhof. Il semble aller de soi pour lui que toute propriété a des aspects. Cependant, s'il n'existe pas de propriétés sans aspects, la théorie des tropes dans son ensemble est tout simplement fautive ! Nul besoin alors de continuer à se questionner sur ses applications en philosophie de l'esprit. Pourquoi se poserait-on la question « en vertu de quoi est-ce efficace causalement ? » à propos d'entités qui n'existent pas ? Pourtant, la théorie des tropes prend clairement position sur la question de l'existence de propriétés sans aspects : les entités postulées – les tropes – sont indivisibles. Elles ne peuvent avoir d'aspects. Qui plus est, il n'est pas nécessaire, dans une telle ontologie, de postuler des niveaux de réalité (même si, rappelons-le, les niveaux de description – notamment dans les théories scientifiques – sont conservés).

En fait, il semble que la critique de Noordhof et la réplique de Robb soient condamnées à se renvoyer l'une à l'autre à l'infini (non sans ironie). Robb affirme que les tropes sont des entités indivisibles, et Noordhof

semble penser que rien n'est moins sûr que l'existence de telles entités. Il semble donc que la façon la plus plausible pour Noordhof de rejeter la théorie des tropes est de s'engager dans la défense de l'affirmation selon laquelle il n'existe pas de propriétés indivisibles. Une telle affirmation conduit tout droit à un enjeu classique en métaphysique : celui de savoir s'il existe des entités dites fondamentales ou non. Bien entendu, nous n'entrerons pas dans le détail de cette discussion ici, mais il est important de noter que la théorie des tropes serait mise à mal si la réponse à cette question était négative. Le théoricien des tropes doit donc prendre parti dans ce débat et défendre l'existence de telles entités primitives, indivisibles. Heil défend d'ailleurs cette position dans un ouvrage paru en 2012 et intitulé *The Universe As We Find It*.

### **3.6 Conclusion**

Nous n'avons pas pu aborder tous les enjeux que soulèvent la théorie des tropes. Toutefois, nous avons montré qu'elle se défendait plutôt bien face aux critiques métaphysiques qui lui sont adressées, et qu'elle permettait une contribution intéressante au débat épistémologique entre réalisme et antiréalisme scientifique. Nous avons conclu qu'elle était donc assez crédible en tant que position ontologique.

Nous avons également montré qu'elle constitue une réponse pertinente au problème de la causalité mentale, en offrant aux réductionnistes une alternative à la position classique défendue par Kim. Nous avons vu que la théorie des tropes est particulièrement efficace pour répondre à l'argument de la réalisabilité multiple. Celui-ci, fréquemment employé par les non-réductionnistes pour faire valoir leur position, ne pose pas autant de problèmes à la théorie des tropes qu'aux autres solutions réductionnistes. Pour finir, nous avons répondu à la critique de Noordhof qui avançait que loin de résoudre le problème de l'exclusion, la théorie des tropes le faisait resurgir au niveau des tropes. Nous en avons conclu que Noordhof semblait en avoir davantage contre la position ontologique elle-même que contre son application à la question de l'efficacité causale du mental. Pour nous, cette dernière est sauvée par la théorie des tropes, et le portrait qui en est peint permet de rendre compte des différentes intuitions qui motivaient la formulation du problème de l'exclusion de façon consistante.

## Conclusion

Depuis l'émergence du fonctionnalisme comme l'une des positions philosophiques les plus en vogue en philosophie de l'esprit à la fin des années 60, peu de progrès ont été faits dans le camp opposé au non-réductionnisme : les éliminativistes sont plus marginaux que jamais, et même les meilleurs efforts de Kim pour réhabiliter un réductionnisme (ou quelque chose qui s'en rapprocherait suffisamment) n'ont pas semblé porter leurs fruits. Le physicalisme est le paradigme dominant, il est vrai. Les dualistes connaissent un certain regain de vie mais ils sont à peine plus nombreux que les éliminativistes. Toutefois, il semble que le conflit qui avait lieu jadis entre dualistes et monistes est aujourd'hui un conflit entre physicalistes réductionnistes et non réductionnistes. La question, qui jusqu'à il y a peu, était « existe-t-il une substance mentale ? » semble être devenue – peut-être face à l'évidence de l'omniprésence du champ lexical du mental dans les discours quotidiens et malgré le développement des neurosciences – « faut-il réduire le mental au physique ? ».

La théorie des tropes offre une solution réductionniste au problème de l'exclusion causale. Bien qu'elle ne fasse pas l'unanimité, et qu'elle ne soit pas une position dominante dans le domaine, elle a de nombreux avantages théoriques dans le cadre de ces débats en philosophie de l'esprit. Elle s'arrime plutôt bien à deux intuitions qui nous tiennent à cœur et dont la conciliation constitue un casse-tête désormais classique : l'efficacité causale du mental d'une part (nos états mentaux se causent entre eux, en plus de causer et d'être causés par des états physiques), et le physicalisme d'autre part. Nous avons montré, dans ce mémoire, que la théorie des tropes fournit au problème corps-esprit une réponse dont la cohérence et l'efficacité méritent qu'on s'y intéresse.

Dans le premier chapitre, nous avons présenté l'argument de l'exclusion, ainsi que certaines des réponses qu'il était possible d'apporter au problème de la consistance des différentes thèses discutées dans celui-ci. Choisir cet argument comme point de départ pour notre discussion semblait pertinent puisqu'il s'agit d'un des arguments les plus discutés dans la littérature. De plus, les versions de la théorie des tropes discutées par Robb, Heil et Whittle répondent directement à celui-ci, il était donc important de l'exposer en détail. Nous en avons présenté la forme, ainsi que les différentes thèses qui le constituent. Ensuite, nous avons montré en quoi ces différentes thèses semblaient inconciliables : d'une part, on souhaite que le mental soit efficace causalement et qu'il ne soit pas détaché du physique afin de ne pas avoir à postuler deux substances distinctes, mais d'autre part, on ne veut pas qu'une propriété mentale et sa base de survenance physique ne soient qu'une seule entité, car cette base de survenance risque d'exclure le rôle causal de la propriété mentale, qui serait alors une propriété épiphénoménale. Une façon bien connue de sortir de cette impasse est le réductionnisme, notamment tel qu'il est défendu par Kim dans ses textes. Cependant, le réductionnisme a des formes multiples, et la position de Kim n'est pas la seule à appartenir à cette famille de solutions. La théorie des tropes, qui nous a intéressés dans ce mémoire, en fait également partie.



Dans le second chapitre, nous avons présenté cette théorie. Il nous a d'abord fallu définir clairement la notion de propriété, et distinguer plusieurs usages qui en sont faits. En effet, nous avons vu que le terme « propriété » était parfois employé pour désigner un élément unique (la couleur particulière de ce livre), mais qu'il pouvait aussi désigner un élément commun à plusieurs objets (la couleur rouge, que possèdent à la fois cette horloge et cette tasse à café). Nous avons alors pu expliquer que selon la théorie des tropes, la réalité était constituée de propriétés individuelles : des particuliers abstraits (les tropes) qui sont les parties constituantes de la réalité, et qui sont également les propriétés responsables de la causalité. Par la suite, nous nous sommes intéressés aux notions de niveau de réalité et de description. Nous avons vu que le monisme des tropes ne postulant qu'un seul type d'entités (les tropes), il n'était pas approprié de conserver cette vision en strates de la réalité dont on retrouve les traces chez plusieurs philosophes – parfois implicitement. Pour la théorie des tropes, il n'existe qu'un seul niveau de réalité : celui constitué par les tropes. Toutefois, ceux-ci tombent sous la coupe de plusieurs prédicats appartenant à des niveaux de description différents. Finalement, nous nous sommes penchés sur la proposition de Whittle, pour qui il est possible de concilier une certaine version de la théorie des tropes avec une analyse fonctionnaliste de la notion de propriété dans son sens universel. Cela nous a permis de définir clairement la notion de classe de ressemblance.

Dans le troisième chapitre, nous avons discuté certains succès explicatifs et de certaines limites de la solution offerte par la théorie des tropes au problème de l'exclusion. Nous avons d'abord montré qu'elle répondait adéquatement à certains problèmes classiques en métaphysique, et qu'elle était également compatible avec une position cohérente en philosophie des sciences : nous avons montré qu'elle était compatible avec l'antiréalisme scientifique, et qu'elle permettait de rendre compte des changements de paradigme dans les domaines scientifiques tout en fournissant un cadre ontologique susceptible de survivre aux changements en question. Ainsi, nous avons pu affirmer qu'il s'agissait bien d'une position ontologique pertinente et légitime. Par la suite, nous avons pu rendre explicite la solution qu'offre la théorie au problème de l'exclusion. Puisqu'une de ses thèses principales est une thèse ontologique selon laquelle la réalité est composée de particuliers abstraits, elle permet de réconcilier les différentes thèses présentées comme inconsistantes dans l'argument de l'exclusion en précisant pour chacune d'entre elles le sens du terme « propriété » qui y est utilisé. Certaines de ces thèses utilisent la notion de propriété dans un sens particulier : elles parlent alors des tropes ; et d'autres l'emploient dans un sens universel : elles parlent alors de classes de ressemblance (selon Heil et Whittle), ou de types (selon Robb 1997). Il s'agit bien d'une solution réductionniste, car toute propriété mentale est identifiée à ce qui est traditionnellement conçu comme sa base de survenance ou de réalisation (une propriété physique) : les deux ne sont en fait qu'un seul et même trope. En fin de chapitre, nous avons parlé du possible retour du problème de l'exclusion au niveau des tropes. Noordhof pense en effet que le problème n'est pas réglé. Pour Robb, toutefois, la critique de Noordhof remet en question un des postulats principaux de la théorie des tropes elle-même : en demandant « *en vertu de quoi*

le trope est-il efficace causalement ? », celui-ci conteste l'idée selon laquelle un trope serait quelque chose d'indivisible, de fondamental. Nous avons conclu qu'une telle critique, pour être valide, devrait se concentrer sur cet aspect de la théorie des tropes et expliquer en quoi il est inadmissible ou incohérent.

Nous pensons, à la suite d'un examen attentif des implications de la théorie des tropes, que la solution qu'elle offre au problème de l'exclusion est sous-estimée. Ses succès explicatifs dans le cadre du problème corps-esprit en font une théorie réductionniste cohérente et n'ayant rien à envier à celle de Kim. En effet, si les états mentaux sont des tropes, ils peuvent tous être des tropes physiques (si nous tenons à conserver la fermeture causale du physique), et cela permet de répondre au problème de l'exclusion en affirmant que le rôle causal d'un trope mental n'est pas exclu par celui d'un trope physique puisqu'il ne s'agit que d'un seul et même trope, tombant sous la coupe de plusieurs classes de ressemblance : l'une mentale et l'autre physique.

Il resterait beaucoup à dire sur la théorie des tropes elle-même, ainsi que sur ses implications dans d'autres champs d'étude de la philosophie. Nous avons par exemple laissé de côté l'analyse historique du concept. Williams – dont nous avons cité les textes afin d'expliquer l'origine de la position de Heil, Robb et Whittle – indique que le terme « trope » est employé par George Santayana, un philosophe né au milieu du XIX<sup>e</sup> siècle. Toutefois, on retrouve des traces de l'idée de particuliers abstraits jusqu'au Moyen-Âge et même dans l'Antiquité. Alain de Libéra a écrit un article paru en 2002 dans la *Revue de Métaphysique et de morale* et intitulé « Des accidents aux tropes. Pierre Abélard », dans lequel il retrace l'histoire du concept jusqu'aux textes de Boèce, d'Avicenne, et même d'Aristote. Également, bien que nous l'ayons défini au second chapitre, le concept d'abstraction n'a pas pu être abordé en profondeur ici, ou en trop grand détail. Il pourrait être pertinent, dans le cadre d'une défense plus systématique de la théorie des tropes, de se demander à quelles critiques il fait face en métaphysique en général, mais également quels enjeux il peut soulever pour l'utilisation de la théorie dans d'autres domaines de la philosophie. Pour ce qui est du problème corps-esprit, mentionnons en terminant que la notion de particulier abstrait semble être tout indiquée pour repenser à neuf, dans un paradigme résolument physicaliste qui s'accorde avec les discussions les plus sérieuses au sein de la communauté scientifique, les entités mentales dont nous avons du mal à saisir la nature tout en ne pouvant nous résoudre à douter qu'elles existent.

## Bibliographie

- Alward, Peter (2008). "Mopes, Dopes, and Tropes: A Critique of the Trope Solution to the Problem of Mental Causation". *Dialogue: Canadian Philosophical Review / Revue canadienne de philosophie*, 47, 1, 53-64.
- Armstrong, David Malet (1999). "The Casual Theory of Properties: Properties according to Shoemaker, Ellis, and Others". *Philosophical Topics*, 26, 1.
- (1989). *A Combinatorial Theory of Possibility*, Cambridge: Cambridge University Press.
- Austin, J. L. (1962). "How to Do Things with Words". *The William James Lectures*, Oxford: Oxford University Press.
- Baker, Alan (2012). "Science-Driven Mathematical Explanation". *Mind*, 121, 482, 243-267.
- Bales, Adam (2020). "Intentions and Instability: A Defence of Causal Decision Theory". *Philosophical Studies*, 177, 3, 793-804.
- Beakley, Brian et Ludlow, Peter (dir.) (1992). *The Philosophy of Mind: Classical Problems/Contemporary Issues*, Cambridge: MIT Press, 2006.
- Beebe, Helen (2017). "Epiphenomenalism for Functionalists". *Making a Difference: Essays on the Philosophy of Causation*, H. Beebe, C. Hitchcock et H. Price (dir.), Oxford: Oxford University Press, 286-306.
- (2004). "Causing and Nothingness". *Causation and Counterfactuals*, J. Collins, N. Hall et L. A. Paul (dir.), Cambridge: MIT Press, 291-308.
- Block, N. J. et Fodor, J. A. (1972). "What Psychological States are Not". *The Philosophical Review*, 81, 2, 159-181.
- Calemi, Francesco F. (dir.) (2016). *Metaphysics and Scientific Realism: Essays in Honour of David Malet Armstrong (Eide Volume 9)*, Berlin: De Gruyter.
- Chakravartty, Anjan (2017a). *Scientific Ontology: Integrating Naturalized Metaphysics and Voluntarist Epistemology*, Oxford: Oxford University Press.
- (2017b) "Scientific Realism". *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.), <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>.

- Churchland, Paul M. (2005). "Functionalism at Forty". *Journal of Philosophy*, 102, 1, 33-50.
- Churchland, Patricia Smith (1988). "Reduction and the Neurobiological Basis of Consciousness". *Consciousness in Contemporary Science*, Anthony J. Marcel et E. Bisiach (dir.), Oxford: Oxford University Press, 273–304.
- Davidson, Donald (1969). "The Individuation of Events". Dans *Essays on Actions and Events (2<sup>nd</sup> Edition)*, Oxford: Clarendon Press, 2001, 163-180.
- Descartes, René (1641). *Méditations Métaphysiques*, Paris : GF Flammarion, 2011.
- Donaldson, John (2021). "Vertical Versus Horizontal: What Is Really at Issue in the Exclusion Problem?". *Synthese*, 198:1381–1396.
- Douglas, Heather E. (2009). *Science, Policy, and the Value-free Ideal*, Pittsburgh: University of Pittsburgh Press.
- Eva, Benjamin et Hartmann, Stephan (2018). "Bayesian Argumentation and the Value of Logical Validity". *Psychological Review*, 125, 5, 806-821.
- Fine, Kit (1975). "Vagueness, Truth and Logic". *Synthese*, 30, 3-4, 265-300.
- Fitzpatrick, William J. (2014). "Skepticism about Naturalizing Normativity: In Defense of Ethical Nonnaturalism". *Res Philosophica*, 91, 4, 559-588.
- Fodor, J. A. (1974). "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)". *Synthese*, 28, 2, 97-115.
- Gibb, Sophie, Lowe, E. J. et Ingthorsson, R. D. (dir.) (2013). *Mental Causation and Ontology*, Oxford: Oxford University Press.
- Gibb, Sophie C. (2014). "Mental Causation". *Analysis*, 74, 2, 327-338.
- Gozzano, Simone et Hill, Christopher S. (dir.) (2012). *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge: Cambridge University Press.
- Gozzano, Simone et Orilia, Francesco (dir.) (2008). *Tropes, Universals and the Philosophy of Mind: Essays at the Boundary of Ontology and Philosophical Psychology*, Hessen: Ontos Verlag.

- Haaparanta, Leila et Koskinen, Heikki J. (dir.) (2012). *Categories of Being: Essays on Metaphysics and Logic*, Oxford: Oxford University Press.
- Hamilton, A. G. (1978). *Logic for Mathematicians (2<sup>nd</sup> Edition)*, Cambridge: Cambridge University Press.
- Heil, John (2012). *The Universe As We Find It*, Oxford: Oxford University Press.
- (2003). *From an Ontological Point of View*, Oxford: Oxford University Press.
- Heil, John et Mele, Alfred (dir.) (1993). *Mental Causation*, Oxford: Oxford University Press.
- Heil, John et Robb, David (2003). "Mental Properties". *American Philosophical Quarterly*, 40, 3, 175-196.
- Hume, David (1748). *An Enquiry Concerning Human Understanding*, Oxford: Oxford University Press, 2007.
- Kim, Jaegwon (2012). "The Very Idea of Token Physicalism". Dans Gozzano et Hill (dir.) (2012), 167-185.
- (2010). *Philosophy of Mind, (3<sup>rd</sup> Edition)*, Cambridge: Westview Press.
- (2008). "Reduction and Reductive Explanation: Is One Possible Without The Other?". *Being Reduced: New Essays on Reduction, Explanation, and Causation*, Howhy, Jakob et Kallestrup, Jesper (dir.), Oxford: Oxford University Press, 93-114.
- (2005). *Physicalism, Or Something Near Enough*, Princeton, N.J.: Princeton University Press.
- (2003). "Blocking Causal Drainage and Other Maintenance Chores with Mental Causation". *Philosophy and Phenomenological Research*, 67, 1, 151-176.
- (1999). "Making Sense of Emergence". *Philosophical Studies*, 95, 1, 3-36.
- (1998). *Mind in a Physical World: An Essay on the Mind–Body Problem and Mental Causation*, Cambridge: MIT Press.
- (1996). *Philosophy of Mind*, Cambridge: Westview Press.
- (1993a). *Supervenience and Mind: Selected Philosophical Essays*, Cambridge: Cambridge University Press.
- (1993b). "The Nonreductivist's Trouble with Mental Causation". Dans Kim (1993), 336-357.

- (1992). "Multiple Realization and the Metaphysics of Reduction", *Philosophy and Phenomenological Research*, 52, 1, 1-26.
- (1989a). "Mechanism, Purpose, and Explanatory Exclusion". Dans Kim (1993), 237-264.
- (1989b). "The Myth of Nonreductive Materialism". Dans Kim (1993), 265-284.
- (1976). "Events as Property Exemplifications". Dans Brand, M. et Walton, D. (dir.), *Action Theory*, Dordrecht: D. Reidel, 310-326.
- (1974). "Noncausal Connections". *Noûs*, 8, 1, 41-52.
- Kim, Jaegwon, Sosa, Ernest et Rosenkrantz, Gary S. (dir.) (2009). *A Companion to Metaphysics (2<sup>nd</sup> Edition)*, Malden, MA: Wiley-Blackwell.
- Kuhlmann, Meinard, Lyre, Holger et Wayne, Andrew (dir.) (2002). *Ontological Aspects of Quantum Field Theory*, Singapour: World Scientific Publishing Company.
- Kuhn, Thomas S. (1996). *The Structure of Scientific Revolutions (3<sup>rd</sup> Edition)*, Chicago: The University of Chicago Press.
- Lewis, David K. (1986a). "Against Structural Universals". *Australasian Journal of Philosophy*, 64, 1, 25-46.
- (1986b). *On the Plurality of Worlds*, Oxford: Blackwell.
- (1980). "Mad Pain and Martian Pain". Dans Block, N. (dir.), *Readings in Philosophy of Psychology*, Vol. 1, Cambridge, MA: Harvard University Press, 216-232.
- (1973). "Causation". *The Journal of Philosophy*, 70, 17, 556-567.
- (1970). "How to Define Theoretical Terms". *The Journal of Philosophy*, 67, 13, 427-446.
- (1966). "An Argument for the Identity Theory". *The Journal of Philosophy*, 63, 1, 17-25.
- Libet, Benjamin (1999). "Do We Have Free Will". *Journal of Consciousness Studies*, 6, 8-9, 47-57.
- List, Christian et Menzies, Peter (2009). "Nonreductive Physicalism and the Limits of the Exclusion Principle". *The Journal of Philosophy*, 106, 9, 475-502.
- Loux, Michael (2006). *Metaphysics: A Contemporary Introduction (3<sup>rd</sup> Edition)*, Londres: Routledge.

- Lowe, Edward Jonathan (2006). "Non-Cartesian Substance Dualism and the Problem of Mental Causation". *Erkenntnis*, 65, 1, 5-23.
- Macdonald, Cynthia et Macdonald, Graham (2010). *Emergence in Mind*, Oxford: Oxford University Press.
- (2006). "The Metaphysics of Mental Causation". *The Journal of Philosophy*, 103, 11, 539-576.
- Magnus, P. D. et Callender, Craig (2004). "Realist Ennui and the Base Rate Fallacy". *Philosophy of Science*, 71, 3, 320-338.
- Malcolm, Norman (1968). "The Conceivability of Mechanism". *Philosophical Review*, 77, 45-72.
- Marmodoro, Anna et Mayr, Erasmus (2019). *Metaphysics: An Introduction to Contemporary Debates and Their History*, Oxford: Oxford University Press.
- Martin, C. B. et Heil, John (1999). "The Ontological Turn". *Midwest Studies in Philosophy*, 23, 1, 34-60.
- McLaughlin, Brian P. et Cohen, Jonathan (dir.) (2007). *Contemporary Debates in Philosophy of Mind*, Malden, MA: Blackwell Publishing.
- Nagel, Ernest (1961). *The Structure of Science*, New York: Harcourt.
- Niiniluoto, Ilkka (2002). *Critical Scientific Realism*, Oxford: Oxford University Press.
- Noordhof, Paul (1998). "Do Tropes Resolve the Problem of Mental Causation?". *Philosophical Quarterly*, 48, 191, 221-226.
- Orilia, Francesco (2006). "Quantum-mechanical Statistics and the Inclusivist Approach to the Nature of Particulars". *Synthese*, 148, 1, 57-77.
- Patterson, Sarah (2005). "Epiphenomenalism and Occasionalism: Problems of Mental Causation, Old and New". *History of Philosophy Quarterly*, 22, 3, 239-257.
- Pearl, Judea (2018). *Causality: Models, Reasoning and Inference (2<sup>nd</sup> Edition)*, Cambridge: Cambridge University Press.
- Pearl, Judea et Mackenzie, Dana (2018). *The Book of Why: The New Science of Cause and Effect*, New York: Basic Books.

- Polger, Thomas W. et Shapiro, Lawrence A. (2016). *The Multiple Realization Book*, Oxford: Oxford University Press.
- Putnam, Hilary (1975). *Philosophical Papers: Volume 2, Mind, Language and Reality*, Cambridge: Cambridge University Press.
- (1967). "The Nature of Mental States". Repris dans *Mind, Language and Reality, Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press, 1975. 429-440.
- Räz, Tim (2014). "On the Application of the Honeycomb Conjecture to the Bee's Honeycomb". *Philosophia Mathematica*, 21, 3, 351-360.
- Richardson, Robert C. (1982). "How Not to Reduce a Functional Psychology". *Philosophy of Science*, 49, 1, 125-137.
- (1979). "Functionalism and Reductionism". *Philosophy of Science*, 46, 4, 533-558.
- Rioux, Catherine (2015). *La théorie de la réalisation par sous-ensembles de Shoemaker et l'exclusion causale* [mémoire de maîtrise en philosophie sous la direction de Renée Bilodeau, Université Laval].
- Robb, David (2013). "The Identity Theory as a Solution to the Exclusion Problem". Dans Gibb, Sophie, Lowe, E. J. et Ingthorsson, R. D. (dir.) (2013).
- (1997). "The Properties of Mental Causation". *The Philosophical Quarterly*, 47, 187, 178-194.
- Russell, Bertrand (1905). "On Denoting". *Mind*, 14, 56, 479-493.
- Sider, Theodore (2003). "What's So Bad about Overdetermination?". *Philosophy and Phenomenological Research*, 67, 719-726.
- Simmons, John A. (2000). *Justification and Legitimacy: Essays on Rights and Obligations*, Cambridge: Cambridge University Press.
- Skinner, Burrhus Frederic (1974). *About Behaviorism*, New York: Vintage.
- Stapp, Henry Pierce (2017). *Quantum Theory and Free Will: How Mental Intentions Translate into Bodily Actions*, New York: Springer.



- (2006). “Quantum Interactive Dualism, II: The Libet and Einstein-Podolsky-Rosen Causal Anomalies”. *Erkenntnis*, 65, 1, 117-142.
- (2006). “Quantum Interactive Dualism: An Alternative to Materialism”. *Zygon*, 41, 3, 599-615.
- Steele, J. Michael (2013). *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge: Cambridge University Press.
- Van Inwagen, Peter (dir.) (1980). *Time and Cause: Essays Presented to Richard Taylor*, Dordrecht: D. Reidel.
- Watson, John Broadus (1930). *Behaviorism*, Chicago: Chicago University Press.
- Whittle, Ann (2008). “A Functionalist Theory of Properties”. *Philosophy and Phenomenological Research*, 77, 1, 59-82.
- (2007). “The Co-Instantiation Thesis”. *Australasian Journal of Philosophy*, 85, 1, 61-79.
- Williams, Donald Cary (1953a). “On the Elements of Being: I”. *Review of Metaphysics*, 7, 1, 3-18.
- (1953b). “On the Elements of Being: II”. *Review of Metaphysics*, 7, 1, 171-192.
- Wilson, Jessica (2005). “Supervenience-Based Formulations of Physicalism”. *Noûs*, 39, 3, 426-459.77, 45-72.
- Wittgenstein, Ludwig (1953). *Philosophische Untersuchungen*. Trad fr. *Recherches philosophiques*, Dastur, F., Élie, M., Gautero, J.-L., Janicaud, D. et Rigal, É, Paris: Gallimard, 2014.
- Woodward, James (2014). “A Functional Account of Causation; or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment)”. *Philosophy of Science*, 81, 5, 691-713.