

2m11.2785.8

Université de Montréal

Les philosophes rêvent-ils en couleur?
L'inversion du spectre et le fonctionnalisme.

par
Frédéric Bouchard
Département de philosophie
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès arts (M.A.)
en philosophie

Août 1999

© Frédéric Bouchard 1999



B
29
U54
2000
v.017

Université de Montréal

L'investissement en recherche et développement
dans les technologies de l'information

par
Frédéric Bourdieu
Département de génie logiciel
Faculté des arts et des sciences

Mémoire présenté à la Faculté des arts et des sciences
en vue de l'obtention du grade de
Maître en sciences (M.Sc.)
en génie logiciel



1999
© Université de Montréal 1999

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé:

Les philosophes rêvent-ils en couleur?
L'inversion du spectre et le fonctionnalisme.

présenté par:

Frédéric Bouchard

a été évalué par un jury composé des personnes suivantes:

M. François Lepage, président-rapporteur

M. Jean-Pierre Marquis, directeur de recherche

M. Daniel Laurier, membre

Mémoire accepté le: 25 novembre 1999

Sommaire

L'inversion du spectre est un argument couramment utilisé pour montrer les limites explicatives du fonctionnalisme. En effet, la possibilité que deux agents aient des expériences qualitatives de la couleur radicalement différentes et que cette différence ne soit pas observable comportementalement, semblent compromettre le fonctionnalisme qui utilise le comportement comme fondation de son cadre théorique.

Le but de ce mémoire est d'examiner la capacité du fonctionnalisme de répondre à ce genre de critique. Après avoir distingué certaines variantes du fonctionnalisme, je présente l'argument métaphysique et pro-fonctionnaliste de Sydney Shoemaker. Favorable à la possibilité de l'inversion du spectre, mais opposé à ce que cette inversion soit nécessairement indétectable, à cause de certains liens avec la réalisation physique des états fonctionnels. Je présente ensuite la critique de Joseph Levine. Ce dernier démontre l'instabilité de l'argument métaphysique de Shoemaker et de son fonctionnalisme supposément conciliant envers la position pro-qualia.

Comme un argument pro-fonctionnaliste métaphysique ne semble pas être efficace, je discute ensuite la tentative totalement a posteriori de Hardin de résoudre ce problème. Il propose en effet un lien intrinsèque entre la perception des couleurs et certaines réactions émotionnelles. Ce lien rendrait impossible une inversion *indétectable*. Levine ne croit pas qu'un tel genre d'argument soit plus efficace que l'argument de Shoemaker. D'une part, une inversion des émotions associées à la perception des couleurs reste possible. D'autre part, l'argument de Hardin, comme tous les arguments matérialistes, serait incapable, selon Levine, d'expliquer *pourquoi* nous avons cette expérience qualitative. J'expose ensuite la réponse de Hardin à ces deux critiques; il montre pourquoi une inversion supérieure n'est pas possible ou du moins pourquoi elle ne pourra pas être indétectable, et pourquoi la critique anti-matérialiste de Levine amène des conséquences sceptiques non-souhaitables. Toutefois, je présente les raisons pour lesquelles la thèse de Hardin ne peut pas être satisfaisante et ce à cause de la nature trop spéculative et empirique de son argument.

Finalement, j'examine enfin le fonctionnalisme téléologique d' Elliot Sober qui cherche à sortir de l'impasse de l'inversion grâce à une théorie s'écartant du physicalisme ou des solutions totalement a posteriori. Sober montre que le problème du fonctionnalisme « classique » est le manque de précision du concept de fonction. En ajoutant la notion de « raison d'être » au concept de fonction Sober croit pouvoir répondre au problème de l'inversion ainsi qu'aux autres problèmes concernant les qualia (comme, par exemple, l'objection anti-matérialiste de Levine). Je veux ainsi montrer que l'inversion du spectre ne sonne pas le glas du fonctionnalisme, mais montre la nécessité de redéfinir le concept de fonction.

Remerciements

Si, comme le dit un certain dicton, cela prend un village pour élever un enfant, cela doit bien prendre un pays pour achever un mémoire. Voici quelques personnes que je tiens à remercier et sans qui ce mémoire ne serait pas ce qu'il est (toutefois, je tiens à préciser que les limites de ce dernier sont de mon entière responsabilité!).

Jean-Pierre Marquis, mon directeur, pour m'avoir fait découvrir des avenues dont j'ignorais même l'existence. La qualité de son suivi et sa disponibilité absolue m'ont permis de me rendre à bon port au moment espéré. Merci beaucoup.

André, mon père, qui a eu la patience de lire et corriger ce mémoire, et Monique, ma mère, qui m'a expliqué que c'était une bonne chose de rêver en couleur. Ils m'ont donné et me donnent toujours l'envie de raffiner ma pensée.

Caroline, ma sœur, qui n'est pas gênée de dire à ses amis ce que je fais dans la vie.

Ma famille et à mes amis qui m'ont obligé, à plusieurs reprises, à expliquer ce qu'est l'inversion du spectre et pourquoi c'est un problème philosophique. Leur intérêt m'a forcé à mieux comprendre mon sujet de maîtrise.

À Jocelyne Doyon, qui m'a montré que détails administratifs et conversations agréables ne s'excluent pas mutuellement.

À François Duchesneau pour m'avoir donné le mot juste au moment où j'en avais besoin.

À Amélie, ma compagne, qui comprend mon sujet de mémoire et qui peut l'expliquer à d'autres (c'est beaucoup!). Pour cela et pour tant d'autres choses, je tiens à la remercier de tout mon cœur.

TABLE DES MATIÈRES

INTRODUCTION.....	1
1 FONCTIONNALISMES.....	7
1.1 CONTEXTE GÉNÉRAL.....	7
1.1.1 <i>L'héritier du béhaviorisme</i>	7
1.1.2 <i>Caractérisation tripartite du fonctionnalisme</i>	9
1.2 RÉALISATION MULTIPLE.....	10
1.2.1 <i>Avantages de la réalisation multiple</i>	10
1.3 FONCTIONNALISME DU SENS COMMUN (OU FONCTIONNALISME ANALYTIQUE).....	14
1.4 FONCTIONNALISME EMPIRIQUE.....	18
2 LE GOUFFRE EXPLICATIF.....	22
2.1 ARGUMENT ÉPISTÉMOLOGIQUE DE LEVINE.....	22
2.1.1 <i>Introduction</i>	22
2.1.2 <i>L'argument de Levine</i>	23
2.1.3 <i>Conclusion</i>	26
2.2 CRITIQUE DE HARDIN.....	27
2.2.1 <i>Introduction</i>	27
2.2.2 <i>Couleur et physicalisme</i>	28
2.2.3 <i>Spéculation et « opponent process theory »</i>	29
2.2.4 <i>Cellule comparatrice et inversion du spectre</i>	33
2.2.5 <i>Conclusion</i>	35
3 INVERSION INTERSUBJECTIVE ET INTRASUBJECTIVE.....	37
3.1 ARGUMENT DE SHOEMAKER EN FAVEUR DE LA POSSIBILITÉ DE L'INVERSION.....	37
3.1.1 <i>Introduction</i>	37
3.1.2 <i>L'argument métaphysique de Shoemaker</i>	37
3.1.3 <i>Inversion intrapersonnelle ou inversion interpersonnelle</i>	40
3.1.4 <i>1^{ère} objection à la possibilité d'inversion du spectre: différence qualitative</i>	43
3.1.5 <i>2^{ème} objection à la possibilité d'inversion du spectre: incompréhension plutôt qu'inversion</i>	45
3.1.6 <i>3^{ème} objection: possibilité concrète d'une inversion</i>	47
3.2 LE NOUVEAU FONCTIONNALISME DE SHOEMAKER.....	48
3.2.1 <i>Parallèles vérificationnistes</i>	48
3.2.2 <i>Réalisation physique</i>	49
3.2.3 <i>Conclusion</i>	51
3.3 CRITIQUE DE LEVINE.....	54
3.3.1 <i>Introduction</i>	54
3.3.2 <i>Confusion entre différents types d'inversion chez Shoemaker</i>	55
3.3.3 <i>Fausse concessions envers la position pro-qualia</i>	60
3.4 CONCLUSION.....	63
4 COOL RED.....	65
4.1 ARGUMENT DE LEVINE.....	65
4.1.1 <i>Introduction</i>	65
4.1.2 <i>L'expérience de pensée telle qu'envisagée par Hardin</i>	65
4.1.3 <i>Le fonctionnalisme chez Hardin</i>	69
4.1.4 <i>D'autres cas d'inversion du spectre et le compromis de Levine</i>	70
4.1.5 <i>Problèmes du matérialisme en général</i>	77
4.2 CRITIQUE DE HARDIN.....	79
4.2.1 <i>Épiphénoménisme des qualia</i>	79

4.2.2 <i>L'implausibilité de l'inversion de Levine</i>	80
4.2.3 <i>Résidu qualitatif</i>	82
4.2.4 <i>Conclusion</i>	83
5 TÉLÉOFONCTIONNALISME	86
4.3 ARGUMENT DE SOBER.....	86
5.1.1 <i>Introduction</i>	86
5.1.2 <i>Fonction biologique plutôt que mathématique</i>	87
5.1.3 <i>Le téléofonctionnalisme et les descriptions a priori</i>	90
5.1.4 <i>Épiphénoménisme des qualia</i>	93
5.1.5 <i>Avantages du téléofonctionnalisme</i>	94
6 CONCLUSION	96
BIBLIOGRAPHIE I	101
BIBLIOGRAPHIE II	105

Introduction

On attribue la première description¹ du problème de l'inversion du spectre à John Locke qui dans «*An Essay concerning Human Understanding*» aurait émis l'hypothèse que «*the same object should produce in several Men's Minds different Ideas at the same time; v.g. the Idea, that a Violet produces in one Man's Mind by his Eyes, were the same that a Marigold produced in another Man's, and vice versa.*»². Cette inversion ne se révélerait pas dans le comportement. Cette indétectabilité a fait de cette expérience de pensée, au début du XX^{ème} siècle, un argument populaire contre les thèses vérificationnistes. On peut donc retrouver divers examens de l'inversion du spectre dans les textes de C.I. Lewis, Moritz Schlick, Hans Reichenbach, John Wisdom, Max Black et J.C.C. Smart³. Toutefois, ce n'est pas la critique contre le vérificationniste que nous examinerons ici, mais plutôt celle contre le fonctionnalisme.

De manière générale, les arguments anti-fonctionnalistes utilisent l'inversion du spectre comme suit: deux agents ont un spectre chromatique inversé l'un par rapport à l'autre. Comme cette inversion est de naissance, l'agent inversé ne sait pas qu'il ne voit pas les couleurs de la même manière que les agents normaux. Ainsi, l'agent normal et l'agent inversé disent tous deux qu'une tomate mûre est rouge, même si l'expérience qualitative de l'agent inversé correspond à l'expérience qualitative de l'agent normal devant le vert. L'anti-fonctionnaliste postule que nous ne serons jamais capables de déceler un cas d'inversion car le comportement de l'agent inversé est le même que celui de l'agent normal. Or le fonctionnaliste décrit la mentalité, comme nous le verrons dans le premier chapitre, en parlant entre autres de comportements. Ainsi, si le comportement se révèle être inadéquat pour rendre compte de la spécificité de l'activité mentale des agents, le fonctionnalisme est incorrect.

¹ Je me base ici sur l'examen offert par Sydney Shoemaker dans «*The Inverted Spectrum*»(1982),p.357.

² Voir Locke, Livre II, chapitre XXXII, section 15.

³ On peut retrouver les références exactes dans le texte de Shoemaker (ibid).

La plupart des arguments utilisent une caractérisation «naïve»⁴ de la couleur: les objets sont tels qu'on les voit, ou qu'on croit les voir. Une critique des différentes thèses relatives au statut physique des couleurs serait peut-être à propos dans un examen de l'inversion du spectre. Toutefois, il ne sera pas fait ici. En effet la question du réalisme ou de l'anti-réalisme de la couleur est un débat à part entière. Le réaliste affirme que les objets ont des propriétés physiques qui correspondent aux couleurs que nous leurs attribuons, alors que l'anti-réaliste ne croit pas que les couleurs correspondent à des propriétés physiques des objets, mais bien des propriétés de notre organisme⁵. Je ne nie pas que cette question a un impact sur la question de l'inversion du spectre: en effet, si les objets ne sont pas vraiment «colorés», comment peut-on dire qu'il y a inversion par rapport à une expérience «objective» des objets? Toutefois, il est possible d'examiner les arguments utilisant l'inversion du spectre en faisant fi de ces considérations. Ainsi, nous accorderons le bénéfice du doute à toutes les thèses examinées ici; le statut de la couleur ne sera pas remis en question, car les arguments ont d'autres éléments indépendants de la résolution de ce problème qui méritent notre attention. De toutes manières, la plupart des arguments restent valables, même si on rejette la conception de la couleur préconisée par l'auteur.

L'inversion est un cas intéressant car il montre la difficulté d'expliquer la mentalité par le comportement. L'inversion est un argument du même type que les cas d'absence de qualia (comme l'argument de la Chine ou des zombies⁶), où l'argument de Jackson à propos de Mary⁷. Le problème général est celui de la possibilité d'une adéquation entre l'expérience phénoménale (les qualia) et le substrat physique sous-jacent. Dans tous ces arguments, on avance que la nature

⁴ C'est ainsi que Shoemaker décrit sa propre conception de la couleur.

⁵ Cette caractérisation est bien sûr très grossière. Je ne veux ici que montrer l'impact potentiel de ce débat sur la question de l'inversion du spectre. Pour une revue introductive, mais poussée de la question du réalisme des couleurs, on peut voir dans Hardin (1991) ou dans les recueils de Byrne et Hilbert (1997)

⁶ Ces arguments sont brièvement examinés dans la section 3.1. L'argument de Shoemaker montrera en fait que l'inversion du spectre est un problème indépendant de celui de l'absence des qualia.

⁷ Le «*Knowledge Argument*» ne sera pas examiné ici, mais pour plus de détails on peut voir, Jackson, F. 1982. *Epiphenomenal qualia*. *Philosophical Quarterly* 32. Plusieurs commentateurs ont continué le débat dont P.M. Churchland 1985.

physique du cerveau ne peut rendre compte de l'expérience phénoménale. Ce sont des critiques générales contre soit le fonctionnalisme, soit le physicalisme, soit le matérialisme. Le cas de l'inversion s'attaque généralement au fonctionnalisme, ce qui rend l'étude du problème mieux circonscrite. Le problème est d'autant plus intéressant que le fonctionnalisme est une théorie qui connaît une certaine popularité, et ce principalement parce qu'elle s'éloigne de la réduction pure et dure à la Churchland (voir entre autres, P.S. Churchland, 1986, et P.M. Churchland, 1988) tout en se déployant dans un cadre matérialiste. Cette théorie semble donc offrir un compromis séduisant pour expliquer la mentalité. Il est important de noter toutefois que l'examen offert ici ne se propose pas d'épuiser la question des théories de l'esprit intéressantes. Plusieurs autres théories donnent d'autres alternatives pertinentes au réductionnisme. Ainsi, une étude de l'externalisme, tel que celui proposé par Putnam ou Burge, pourrait jeter une lumière intéressante sur la question. Toutefois, ces théories ne seront pas examinées ici.

Reste à savoir comment traiter de la question de l'inversion. Il y a bien sûr le problème de l'inversion des qualia en général, mais l'inversion du spectre offre un cas plus instructif. Comme le souligne Shoemaker (1980, p.256), le cas de l'inversion du vert est un cas plus facile à étudier que le cas d'une inversion de la douleur, par exemple; d'une part, il est difficile de croire qu'une inversion de la douleur ne provoquerait aucun changement comportemental, alors que l'inversion de deux couleurs semblent être plus difficile à déceler. Or c'est cette difficulté qui rend l'inversion du spectre un argument valable contre le fonctionnalisme. L'inversion des qualia n'est donc pas un contre-argument aussi intéressant que l'inversion du spectre. D'autre part, il est difficile de concevoir une distinction entre la douleur et l'état qualitatif de la douleur, alors que l'on peut plus facilement concevoir une distinction entre voir du vert et avoir l'expérience phénoménale du vert. Cette distinction est importante car l'argument anti-fonctionnaliste consiste justement à dire que les états fonctionnels sont «totalement» détachés (bien sûr, l'écart entre le fonctionnaliste et le phénoménal diffère d'un auteur à l'autre) des états qualitatifs. Le cas de la douleur n'est donc pas aussi significatif car on a de la difficulté à

séparer les deux états. L'inversion du spectre est donc un cas privilégié car il offre un argument intéressant contre le fonctionnalisme tout en étant assez circonscrit pour que son attaque soit pertinente. De plus l'inversion du spectre est définie de sorte à être plus convaincante que l'inversion des qualia en général et, argument non-négligeable, elle semble être une possibilité empirique⁸, ce qui rend cette expérience de pensée et le débat qu'elle engendre plus faciles à accepter.

Avant de voir comment l'inversion du spectre attaque le fonctionnalisme, il est important de bien comprendre ce qu'est le fonctionnalisme et à quoi correspondent ses différentes variantes. C'est ce qui est fait dans la première partie de ce mémoire. J'examine la caractérisation du fonctionnalisme offerte par Braddon-Mitchell et Jackson et surtout la dichotomie qu'ils soulignent entre le fonctionnalisme du sens commun (aussi appelé fonctionnalisme analytique) et le fonctionnalisme empirique. Cet examen révélera l'importance de bien qualifier le fonctionnalisme dont on parle. Ainsi, comme nous le verrons tout au long de ce mémoire, la plupart des commentateurs ne semblent pas distinguer les différentes formes de fonctionnalisme. Ceci diminuera l'efficacité de leur argument.

Après avoir défini le fonctionnalisme, il s'agit de voir s'il est possible de rejeter le fonctionnalisme sans tenir compte de ses variantes. C'est ce qui sera fait dans la seconde partie de ce mémoire. Ainsi, nous examinerons l'argument de Levine contre le fonctionnalisme. Levine espère montrer, grâce à l'inversion du spectre, que toute thèse fonctionnaliste (qu'elle soit empirique ou analytique) est vouée à l'échec de même que toute thèse matérialiste en général et ce à cause du gouffre explicatif entre le physique et le qualitatif. J'espère montrer que cette attaque n'est pas aussi décisive que ne le souhaite Levine et qu'il est nécessaire de déterminer quelle est la thèse fonctionnaliste que l'on désire attaquer, car les particularités des variantes fonctionnalistes rendent difficile le rejet en bloc du

⁸ Je ne me pencherai pas ici sur la possibilité empirique de l'inversion du spectre, car comme je le dirai à d'autres reprises dans ce mémoire, même la preuve éventuelle de son impossibilité pratique n'attaque pas sa valeur théorique. En effet, même si on montrait que l'inversion du spectre n'est pas possible chez les humains, il faudrait rendre compte de la possibilité d'une telle inversion chez d'autres espèces. Toutefois pour un argument empirique en faveur de la possibilité de l'inversion du spectre, on peut examiner le texte de Nida-Rümelin, 1996.

fonctionnalisme. L'argument de Levine est suivi d'une réplique de Hardin qui cherchera à défendre le matérialisme en se servant de l'inversion du spectre et d'une description empirique particulière de la perception chromatique. Il est intéressant de voir que Hardin, lui non plus, ne distingue pas les variantes fonctionnalistes et ne croit pas que sa théorie puisse être accommodée par aucune de celles-ci, ce qui, comme je souhaite le montrer, reste à démontrer.

La réplique foncièrement empirique de Hardin montrant certaines faiblesses, nous examinerons ensuite la possibilité de défendre le fonctionnalisme de manière *a priori*. C'est ce qui sera fait dans la troisième partie du mémoire: nous examinerons la théorie de Shoemaker qui cherche, non pas à fonctionnaliser les qualia, mais bien à fonctionnaliser la classe des qualia. Il essaie de rendre compte du cas de l'inversion du spectre dans un cadre matérialiste de manière générale, mais il essaie aussi de fournir une nouvelle manière d'envisager le fonctionnalisme pour accommoder les qualia sans les éliminer. Cet examen sera suivi de la critique de Levine qui montrera que ce «nouveau» fonctionnalisme n'est pas aussi conciliant que ne le souhaite Shoemaker car il enlève toute pertinence aux qualia.

Dans la quatrième partie du mémoire, nous reprendrons la querelle entre Levine et Hardin qui essaient, presque dix ans après leur première confrontation, de revenir à la question de l'inversion du spectre. Comme la défense *a priori* du fonctionnalisme, telle que présentée par Shoemaker, n'est pas concluante, nous en reviendrons donc à des considérations plus empiriques. Cette fois, Levine et Hardin offrent tous deux un examen du lien potentiel entre les couleurs et les émotions et comment une éventuelle association intrinsèque entre la perception de couleurs et une réaction émotionnelle face à ces couleurs pourrait affecter le débat. Levine en vient à déplacer l'inversion au-delà des couleurs, pour dire qu'une inversion des émotions associées aux couleurs serait possible. Hardin réplique qu'une telle inversion est soit impossible soit contre-productive pour le pro-qualia.

Enfin, dans la cinquième partie, nous examinerons une manière de caractériser le fonctionnalisme qui permettrait peut-être de sortir de l'impasse. Ainsi le téléofonctionnalisme, en réintégrant la notion de «raison d'être» dans le concept

de fonction, pourrait peut-être balayer le problème de l'inversion du spectre. Plusieurs variantes du téléofonctionnalisme existent, mais nous verrons celle de Sober (1990) qui a l'avantage d'offrir de manière synthétique une réponse à la plupart des défis lancés de Levine⁹. De plus Sober, en tant que philosophe de la biologie, jette un regard différent, et peut-être plus éclairé, sur le concept de «fonction».

Mon intention générale dans ce mémoire est de montrer que le fonctionnalisme, du moins dans ses caractérisations usuelles, est mal défini, cette faiblesse étant évidente dans le cas de l'inversion du spectre. Toutefois, j'espère aussi montrer que les arguments anti-fonctionnalistes ne sont pas décisifs. Ainsi, je ne crois pas que l'inversion du spectre doit nécessairement montrer l'échec du fonctionnalisme mais bien la nécessité de redéfinir le fonctionnalisme. Comme on peut le voir à la fin du mémoire, le téléofonctionnalisme semble à ce stade-ci une voie prometteuse.

⁹ Comme on peut facilement le remarquer, mon mémoire utilise beaucoup les textes de Levine. Il n'est pas le seul «pro-qualia» à s'être attaqué à la question de l'inversion du spectre, mais il est celui qui s'y est le plus intéressé et qui a généré le plus de critiques (dont celles soulevées dans ce mémoire). C'est pourquoi nous avons limité la description de la position «pro-qualia» à celle de Levine. Ceci nous offre un point de repère clair et stable dans notre critique de cette position (de plus, elle est relativement représentative des autres critiques anti-fonctionnalistes utilisant l'inversion du spectre).

1 Fonctionnalismes

1.1 Contexte général

1.1.1 L'héritier du béhaviorisme

Avant de montrer comment le fonctionnalisme peut répondre à l'argument de l'inversion du spectre, il est d'abord nécessaire de présenter le fonctionnalisme et ses variantes contemporaines. Il ne s'agira pas ici de présenter les sources historiques de chacune de ses souches¹⁰, mais bien de décrire le fonctionnalisme dans sa généralité pour ensuite présenter quelques tangentes qu'il a engendrées en réponse aux critiques du type de l'inversion du spectre. Ce survol nous permettra de mieux apprécier la cible réelle des arguments pro-qualia, et de mieux caractériser les réponses pro-fonctionnalistes qui suivirent.

Le point de départ sera la caractérisation offerte par David Braddon-Mitchell et Frank Jackson dans *Philosophy of Mind and Cognition*. Leur étude du fonctionnalisme illustrera les différences entre le *common-sense functionalism* (aussi appelé fonctionnalisme analytique) et les différentes sortes de *empirical functionalism*.

Le fonctionnalisme est souvent décrit comme étant l'héritier de la thèse maintes fois discréditée qu'est le béhaviorisme¹¹. Cette thèse n'utilise pas le concept d'état mental, mais seulement celui de comportement (et, dans certaines variantes du béhaviorisme, des dispositions à avoir certains comportements). Sans rentrer dans les détails, le béhaviorisme exhibe rapidement ses limites explicatives; bien qu'il semble à première vue satisfaisant pour rendre compte des comportements des autres agents, il n'a pas les ressources explicatives nécessaires (et ne cherche pas à les avoir) pour rendre compte de la vie mentale à la première personne. Ainsi, une introspection rapide révèle immédiatement que notre vie mentale ne peut pas être réduite à des comportements ou à des dispositions. L'objection évidente faite au béhaviorisme est qu'il ne peut pas rendre compte adéquatement de la douleur. En

¹⁰ Une telle entreprise sera d'autant plus ardue que chaque auteur imprime son mouvement distinct au «fonctionnalisme».

¹¹ La première grande «défaite» du béhaviorisme a été causée par Chomsky et ses commentaires quant à la théorie de l'apprentissage du langage.

effet, en misant l'explication du mental exclusivement sur le comportement, l'agent extérieur serait une source plus sûre d'information sur la vie mentale en sa qualité d'observateur extérieur des comportements. Ceci semble absurde dans le cas de la douleur où l'agent est davantage en mesure de dire s'il souffre, même s'il n'exhibe pas de comportements associés. Il est important de noter que le béhaviorisme n'a jamais eu la prétention de rendre compte d'une quelconque mentalité, mais seulement du comportement. Cette limitation du domaine de recherche n'est pas erronée en soi, mais elle ne semble pas en mesure de rendre compte pleinement du comportement humain. En effet, elle ne rend pas justice à l'intuition que nous avons concernant l'efficacité du mental sur le comportement. De plus, même si le béhavioriste cherchait à se débarrasser de la douleur en tant qu'état mental, il doit expliquer comment la douleur semble modifier le comportement qui, lui, fait partie de son domaine de recherche.

Une théorie de l'esprit qui ne peut pas rendre compte de la douleur doit intuitivement être rejetée¹². De manière générale, le béhaviorisme ne veut pas expliquer adéquatement l'introspection ou la raison d'être de l'expérience phénoménale. En ne gardant que les comportements ou les dispositions, le béhaviorisme soustrait tout rôle causal aux états mentaux. L'intention ou la méthode du béhaviorisme pouvait sembler louable (et elle est d'ailleurs encore utilisée en psychologie et dans les sciences cognitives en général), mais en ne maintenant que le caractère observable à la troisième personne, il commet les mêmes erreurs et manifeste les mêmes limites que le vérificationnisme.

Le problème plus général du béhaviorisme est qu'il ne peut pas réduire les états mentaux de la manière dont il voudrait le faire. Les comportements ne sont pas assez distincts pour rendre compte de tous les états mentaux qu'on croit avoir¹³. La relation entre comportement et état mental ne semble pas être une relation d'identité

¹² Jackson rapporte la blague classique illustrant les lacunes du béhaviorisme: « *The weaker of these [objections] is the one contained in the joke about two behaviourists meeting on the street. One says to the other, « You feel fine, how do I feel? » The joke assumes that behaviourists think that we have better access to the mental states of others than we do to our own.* » p.33

¹³ Le problème est particulièrement frappant lors de la description de la douleur. Devant certains comportements nous sommes prêts à affirmer que l'agent souffre, mais nous ne pouvons pas vraiment établir la gravité de sa douleur seulement en fonction de son comportement.

stricte, mais bien une relation d'association plus contingente¹⁴.

1.1.2 Caractérisation tripartite du fonctionnalisme

Le fonctionnalisme sera la réponse naturelle à ce genre d'objection en introduisant les états mentaux dans la chaîne explicative. Il essaie de remplir la boîte noire béhavioriste qui médiatise les *input* et les *output*, avec quelque chose qui laisse une certaine place à ce que l'intuition populaire qualifie d'activité mentale¹⁵. Ainsi, la vie mentale devra maintenant être expliquée par la triade suivante: il y a les *input* qui concernent les conditions d'apparition de certains états mentaux, les interactions internes entre les états mentaux (ce que le béhaviorisme négligeait de manière flagrante), et les *output* qui montrent quels états mentaux provoquent quels comportements particuliers. Alors que le béhaviorisme décrivait la vie mentale en termes de réactions à des stimuli, le fonctionnalisme réintroduit l'idée de causalité mentale comme intermédiaire entre l'environnement et le comportement. Cette idée de causalité est centrale car les états mentaux sont identifiés selon leur cause et leur effet sur le comportement ou sur d'autres états mentaux. Ce genre d'explication semble plus intuitif que le béhaviorisme car il redonne une certaine valeur à l'introspection en tant qu'elle semble être en mesure de pointer vers certains éléments de la chaîne causale étudiée (ce que le béhaviorisme ne reconnaissait pas). Cette description tripartite de la vie mentale est le point commun de toute thèse fonctionnaliste. Les divergences entre les diverses variantes fonctionnalistes apparaissent immédiatement après ce point d'entente. Ainsi, la réalisation multiple qui semble être un autre point d'entente de toutes les thèses fonctionnalistes mérite d'être mieux qualifiée, en regard de la tangente fonctionnaliste adoptée. Comme nous le verrons, le fonctionnaliste doit nécessairement adopter la réalisation

¹⁴ « *The problem for behaviourism about the mind is that there is no such thing as the distinctive behaviour associated with any mental state in itself, there just is no consistent behaviour that will typically follow any given mental state unless there is some sort of outside interference.* » Braddon-Mitchell et Jackson p.38

¹⁵ Un neurologue dirait que la manière « naturelle » de remplir la boîte noire est de donner une cartographie des neurones, mais ce genre d'explication semble tout à fait contre-intuitif pour la plupart des profanes. Ainsi, l'explication fonctionnaliste semble plus séduisante car elle essaie de concilier nos intuitions matérialistes (l'esprit C'EST le cerveau) et nos intuitions dualistes (notre esprit N'EST PAS QUE neurones).

multiple, toutefois le degré de la « multiplicité » de ces réalisations variera énormément d'une thèse à l'autre.

1.2 Réalisation multiple

1.2.1 Avantages de la réalisation multiple

La thèse de la réalisation multiple est un des éléments les plus séduisants du fonctionnalisme. Elle affirme que la nature exacte du substrat physique du système fonctionnaliste désiré n'est pas figée par la définition fonctionnelle. Ainsi, pour juger de l'isomorphie fonctionnelle, il n'importe pas de savoir si tel réseau neuronal est exactement identique à tel autre réseau neuronal, mais bien de savoir si ces deux réseaux occupent la même fonction (ont la même cause et le même effet) dans l'explication fonctionnelle dans laquelle ils s'inscrivent. Rejeter la réalisation multiple revient à refuser *a priori* une conscience ou une expérience phénoménale analogue à la nôtre, à tout organisme qui pourrait (du moins intuitivement) avoir ces éléments communs avec nous mais qui aurait une constitution différente de la nôtre. L'exemple du Martien hypothétique va dans ce sens. Nous pouvons imaginer des êtres qui, par exemple, seraient des organismes à base de silicone plutôt que de carbone, mais qui exhiberaient les mêmes comportements que nous, révéleraient des états d'âme analogues aux nôtres, etc. Ainsi, la description fonctionnaliste permet théoriquement¹⁶ d'attribuer à ces êtres la même conscience et la même expérience phénoménale que la nôtre si et seulement si leurs inputs, leurs interactions entre les états mentaux et les outputs sont identiques aux nôtres. D'une manière plus restreinte, la réalisation multiple permet d'expliquer la vie mentale de deux agents qui seraient structurellement différents tout en étant fonctionnellement isomorphes.

Braddon-Mitchell et Jackson (p.43) identifient cinq raisons qui rendent la réalisation multiple (et, par la bande, le fonctionnalisme) attrayante¹⁷. Un examen détaillé de ces raisons est nécessaire car plusieurs arguments « anti-inversion »

¹⁶ Ce ne sont pas tous les fonctionnalistes qui accepteront ceci, mais une réalisation si radicalement différente de la nôtre illustre bien le point. En fait, le degré de réalisation multiple sera souvent plus limité par les fonctionnalistes.

¹⁷ Je regrouperai (3) et (4) dans la même catégorie car ils relèvent tous deux de la différence des cerveaux. Dans le cas de (3) cette différence est interpersonnelle alors que dans (4) elle est

essaieront d'utiliser la réalisation multiple pour tenter de sauver le fonctionnalisme.

(1) Dans la vie de tous les jours nous décrivons l'action des agents et leurs comportements sans nous préoccuper ou sans même être en mesure de comprendre la nature exacte de ce qui les réalise. Malheureusement cet argument ne prouve pas la valeur de ce genre de prédiction. Comme le montreront à répétition les matérialistes éliminativistes, le sens commun s'est déjà trompé sur ce qu'il prenait auparavant pour évident. Le sens commun ne semble donc pas suffisant pour garantir la validité de la thèse de la réalisation multiple. Bien qu'attrayant, cet argument de Braddon-Mitchell et Jackson n'est donc pas concluant.

(2) L'exemple du Martien illustre que nous ne devrions pas être « chauvins » concernant l'esprit. Le rejet de la réalisation multiple implique le rejet de conscience non-humaine ce qui semble être, à ce stade-ci, non justifié. Cet argument faisant appel à la prudence face à la caractérisation de la mentalité est, selon moi, le problème le plus sérieux de ceux qui rejettent la réalisation multiple comme le font certains matérialistes éliminativistes. Le rejet éliminativiste de la mentalité non-humaine peut être considéré de deux manières: soit comme un problème épistémologique (un peu comme le rejet davidsonien de la mentalité animale face à notre incapacité de rendre compte des comportements animaux, ce qui rend inapplicable le principe de charité), soit comme un rejet métaphysique de la *possibilité* d'une mentalité non-humaine. Dans les deux cas, la conclusion est hâtive: d'un côté, le rejet métaphysique semble être davantage une pétition de principe qu'un argument sérieux contre des mentalités parallèles, et par ailleurs le rejet épistémologique (qui peut être un simple scepticisme local quant à des « esprits »¹⁸ non-humains) semble devoir déborder vers un scepticisme quant aux autres esprits (humains) et *a fortiori* vers un scepticisme global. Plutôt que d'emprunter cette voie radicale, l'éliminativiste devrait laisser le fonctionnalisme faire ses preuves. Ainsi autant les éliminativistes nous demandent de leur laisser la chance de montrer que

intrapersonnelle.

¹⁸ Il est important de souligner que l'éliminativiste ne se servira pas de termes comme « mentalité » ou « esprit » car ce sont justement le genre de termes que son type d'explication matérialiste cherche à éliminer. Toutefois, le nouveau vocabulaire servant à expliquer l'action humaine n'ayant pas encore été établi, l'utilisation des termes classiques semble ici encore appropriée.

l'on peut éliminer la psychologie, autant ils ne devraient pas rejeter *a priori* la possibilité de la réalisation multiple, et donc d'une explication fonctionnaliste. Provisoirement devant le choix entre réduction et élimination, la réduction semble être le choix le plus prudent en donnant son dû à la neuro-anatomie sans discréditer totalement la compréhension offerte par le sens commun quant à la mentalité.

(3) L'organisation structurelle des cerveaux humains varie de manière importante d'une personne à une autre. Ainsi alors que le langage occupera une zone x chez l'agent A, il peut occuper une zone y chez l'agent B. Cette polyvalence neuronale est un indice frappant de la réalisation multiple. Cet argument présenté par Braddon-Mitchell et Jackson, bien que convainquant, est plus modeste qu'il n'apparaît à première vue. Ainsi, même s'il semble un point fort en faveur de la réalisation multiple en général, cet argument ne fait en réalité que renforcer la possibilité de réalisation multiple *chez les humains*. Ce n'est pas parce que deux cerveaux humains sont différents et qu'ils *font* la même chose que l'on peut conclure qu'un cerveau radicalement différent (le Martien par exemple) peut soutenir une organisation fonctionnelle analogue. Ainsi cet argument, bien que probant, n'est pas aussi généreux ou « non-chauvin » que ne le voudraient les fonctionnalistes de la première heure. En fait, certains fonctionnalistes empiriques (nous reviendrons bientôt sur cette qualification distincte) voient dans ce genre de critique une certaine limite que devront respecter les explications fonctionnalistes. Ce ne sont donc pas tous les fonctionnalistes qui affirment que l'organisation fonctionnelle ne dépend pas du tout du substrat qui la supporte.

Braddon-Mitchell et Jackson présentent aussi le cas des agents qui se retrouvent avec une partie de leur cerveau endommagée à la suite d'un accident cérébro-vasculaire (infarctus, anévrisme, embolie, etc.) et dont une autre zone du cerveau prend la relève pour occuper la fonction opérée précédemment par la zone maintenant altérée. L'argument est le même pour prouver la validité de la réalisation multiple. Toutefois la portée de l'argument est d'autant plus limitée qu'elle cherche ici à déduire injustement d'un cas de réalisation multiple *intrapersonnelle* non seulement une réalisation multiple *interpersonnelle*, mais en plus une réalisation

multiple « interspécifique » (entre espèces).

(4) ((5) dans le texte original¹⁹) Il ne semble pas farfelu d'imaginer le remplacement progressif de certaines sections du cerveau par des puces de silicones. Si on peut admettre le remplacement avec succès (ce qui veut dire que les mêmes fonctions sont jouées par la puce) d'une section du cerveau, on peut imaginer un « cerveau » totalement constitué de puces. L'organisation fonctionnelle resterait la même malgré une réalisation physique différente. Il est à noter que ce genre d'argument fait fi de la possibilité que le tout est supérieur à la somme des parties. Ainsi il se peut que le remplacement d'une ou deux sections ne dérangera pas le tout, mais il est loin d'être évident que le remplacement complet de neurones par des puces assure les mêmes fonctions²⁰. Nous allons provisoirement accepter cette expérience de pensée comme étant du moins concevable, mais il serait pertinent d'examiner la possibilité que l'existence de propriétés émergentes attaque la validité de cette expérience de pensée (cependant cet examen ne sera pas développé dans cet exposé).

Nos observations et nos intuitions semblent indiquer un modularisme du cerveau et une flexibilité de ces composantes et de leurs organisations. La réalisation multiple semble la plus en mesure de rendre compte de ces aspects, et le fonctionnalisme embrasse la réalisation multiple d'une manière inégalée parmi les théories matérialistes de l'esprit. Restera bien sûr à savoir quelle sera la portée de cette réalisation multiple. En fait, la définition du domaine des réalisations parallèles qui seront acceptées comme étant fonctionnellement isomorphes à notre esprit, établira l'espoir ou l'échec du fonctionnalisme. Comme j'espère le montrer, ce domaine devra être relativement restreint, ce qui éliminera d'office le fonctionnalisme du sens commun (fonctionnalisme analytique), mais qui protégera certaines variantes du fonctionnalisme empirique des critiques basées sur des arguments comme l'inversion du spectre.

Braddon-Mitchell et Jackson identifient sept variantes du fonctionnalisme. Ils classifient ses variantes en deux grandes familles. Il y aurait le fonctionnalisme du

¹⁹ Braddon-Mitchell et Jackson voyaient dans le cas des accidents cérébro-vasculaires, un argument distinct de 3), mais la faiblesse de l'argument me pousse à le joindre à 3) plutôt que de le traiter comme un argument autonome.

sens commun (ou fonctionnalisme analytique) et six sortes de fonctionnalismes empiriques²¹. Sans entrer dans les détails, qui ne sont pas utiles à notre propos, on peut dire que les variantes empiriques se distinguent par le rôle qu'elles attribuent au sens commun et par la caractérisation qu'elles offrent des inputs et des outputs (pour certaines variantes, ils correspondent à des états physiques, alors que pour d'autres ils correspondent à des relations plus abstraites avec l'environnement). Cette classification m'apparaît incomplète car elle semble faire fi des particularités des variantes plus téléologiques du fonctionnalisme (comme le fonctionnalisme homonculariste de William Lycan ou le téléofonctionnalisme de Sober que nous examinerons brièvement dans la dernière section de ce mémoire). Il n'est pas nécessaire, pour notre propos, d'examiner en détails la caractérisation que Braddon-Mitchell et Jackson offrent du fonctionnalisme. Un examen sommaire des différentes familles fonctionnalistes sera suffisant pour montrer que l'argument de l'inversion du spectre ne peut pas attaquer tous les fonctionnalismes de la même manière et qu'ainsi, il ne constitue pas une raison suffisante pour rejeter *tous* les fonctionnalismes *a priori*.

1.3 Fonctionnalisme du sens commun (ou fonctionnalisme analytique)

Dans la plupart des arguments utilisant l'inversion du spectre, le fonctionnalisme analytique est la cible visée (en fait, les autres sortes de fonctionnalismes sont souvent injustement identifiées à lui). On qualifie souvent ce fonctionnalisme de « pur » car il est la forme de fonctionnalisme qui semble la moins attachée à définir un substrat physique nécessaire ou une limite stricte au domaine des réalisations multiples acceptables²². C'est la forme de fonctionnalisme privilégiée par Braddon-Mitchell et Jackson, mais c'est celle qui semble la moins instructive quant à la vie mentale des agents.

Tout fonctionnalisme doit déterminer quelles sont les fonctions à expliquer. Chaque élément à étudier joue plusieurs rôles, mais le fonctionnalisme assume que

²⁰ Ce qui serait le cas si le cerveau était un système «chaotique».

²¹ Pour un excellent tableau récapitulatif des différentes variantes fonctionnalistes, voir Braddon-Mitchell et Jackson p.88-89. Ce tableau nous servira grandement à établir les différences et les ressemblances entre les différentes formes de fonctionnalismes.

seuls certains rôles sont pertinents. Ainsi c'est un peu comme si on disait qu'un magnétoscope donne l'heure (et que c'est pratique), mais si on le voit comme une grosse horloge rectangulaire, on ne peut pas comprendre pourquoi il a des têtes d'enregistrement à l'intérieur. Le fonctionnalisme ne peut pas être décrit seulement en termes d'une explication causale, car plusieurs effets, sans être accidentels, ne concernent pas le rôle « premier » de l'élément étudié. Tout fonctionnalisme devra se munir de moyens d'identifier le rôle premier, la fonction « réelle » des constituantes de son système. Le fonctionnalisme analytique se servira du sens commun pour identifier quelles sont les fonctions « importantes ». Le sens commun établira la signification des termes mentaux²³ qui sont inclus dans le système étudié. Ce critère « communautaire » a l'avantage, comme le soulignent Braddon-Mitchell et Jackson (p.46), de se baser sur le lien intuitif que nous établissons entre l'esprit et le comportement. Ainsi, les termes mentaux définis par le sens commun cherchent habituellement à rendre compte de ce qui est observable pour l'observateur extérieur, soit le comportement. Toutefois, en incluant les termes mentaux dans la chaîne causale et en faisant des définitions de ces termes des conventions tacites acceptées par la majorité, le fonctionnalisme analytique semble être beaucoup plus plausible que le béhaviorisme qui lui aussi se base sur le comportement.

Toutefois, et Braddon-Mitchell et Jackson en sont conscients malgré leur endossement du fonctionnalisme analytique, le sens commun montre rapidement ses limites. Ainsi à partir de quel moment une définition est-elle assez partagée pour faire partie d'une explication fonctionnelle? Le fonctionnalisme analytique tombe de Charybde en Scylla. Si le critère de sélection des fonctions est trop strict, plusieurs fonctions seront rejetées *a priori* parce que jugées non pertinentes par le sens commun. Si le critère de sélection des fonctions est trop généreux, on attribuera injustement des définitions fonctionnelles à des unités (surtout les termes mentaux) qui ne jouent pas vraiment un rôle fonctionnel pertinent.

²² Ce fonctionnalisme est aussi appelé, dans la littérature, « fonctionnalisme classique ».

²³ Comme le soulignent Braddon-Mitchell et Jackson (p.46) c'est pour cette raison que l'on qualifie ce fonctionnalisme d'analytique car il cherche à donner une explication fonctionnelle de l'esprit humain (et donc une explication d'une partie du monde réel) grâce à l'élucidation des concepts d'états mentaux.

Le fonctionnalisme analytique est condamné à utiliser des définitions disjonctives (*cluster concepts*). Comme les définitions dépendent du sens commun, elles ne sont pas uniques et strictes. Ce qui semble évident pour un individu peut sembler étrange à un autre. Ainsi, la seule manière d'établir des définitions sera de constituer des banques d'attributs dont la fonction devra instancier un certain nombre de propriétés. Par exemple, le concept de plaisir pourrait être une disjonction de 1500 propriétés; mais pour dire qu'un état correspond au plaisir, il ne devrait pas nécessairement présenter 1500 propriétés, mais peut-être seulement 900. Un autre état pourrait aussi être défini en tant que plaisir s'il possède aussi 900 des 1500 propriétés, même si ce ne sont pas les mêmes 900 propriétés que le premier état. Cette description du critère des fonctionnalistes analytiques révèle immédiatement les limites de cette position. Le fonctionnaliste analytique ne peut pas parler en termes de nombres de propriétés: le sens commun n'offre pas des définitions aussi strictes. Braddon-Mitchell et Jackson ne sont pas prêts à dire qu'une définition d'un terme mental doit avoir 1500 ou 100000 de propriétés et *a fortiori* ils ne sont pas prêts à dire combien de propriétés doivent être instanciées par un état particulier. En fait le langage qu'ils utilisent est beaucoup plus vague: « *But nothing in this list is sacrosanct (...) What matters is that enough of the list is satisfied or near enough satisfied, and what counts as enough may itself be a vague matter, and may change with time. But enough of the list must be satisfied or near enough satisfied.* » (Braddon-Mitchell et Jackson p.47-48).

D'autre part, le fonctionnalisme analytique ne semble pas être en mesure de dire ce qui sera suffisant pour remplir les conditions de cette «liste» car le sens commun ne pourra arriver au consensus nécessaire pour l'élaboration de critères stricts. Braddon-Mitchell et Jackson sont conscients du problème²⁴ et croient que les définitions doivent avoir comme fondation nos connaissances implicites et tacites quant à la mentalité. L'efficacité de notre pouvoir prédictif quant aux actions des autres agents serait un argument en faveur de cette approche. Nous prédisons très souvent avec succès l'action des autres agents en les considérant comme des agents

²⁴ « *the hardest [question] is about how, precisely, to specify the raft of common-sense opinions about*

possédant des états mentaux, et ces états mentaux sont définis par notre sens commun. Ce genre d'argument est toutefois problématique, car en se basant sur le pouvoir prédictif il est sujet à un renversement facile d'une théorie montrant un plus grand pouvoir prédictif. Ainsi, si une autre théorie démontrait un plus grand pouvoir prédictif sans se baser sur le sens commun, le fonctionnalisme analytique devrait accepter la défaite. Plus important encore, le fonctionnalisme analytique ne peut expliquer de manière satisfaisante pourquoi nous ne sommes pas capables d'offrir des définitions strictes. Braddon-Mitchell et Jackson tentent de fournir une explication: « *True, we cannot write down the rules we follow (although we can give some rough examples), and from an evolutionary point of view, that is to be expected. Making predictions, rather than theorizing about how to make them is what matters for survival.* » (p.56).

Cette explication est insatisfaisante car le haut degré de réalisation multiple accepté par le fonctionnalisme analytique semble empêcher a priori tout recours à une thèse évolutionniste. Ainsi, si nous acceptons qu'un ordinateur ait une conscience, peut-on invoquer la théorie de l'évolution pour dire que l'ordinateur n'a pas de règles strictes pour nous attribuer la conscience? J'assume ici, peut-être à tort, que l'ordinateur ne fait pas partie de l'histoire darwinienne. C'est un point discutable. Toutefois, il semble étrange d'évoquer la théorie de l'évolution pour défendre le flou d'un critère fonctionnaliste (celui du sens commun) alors que l'on voudrait pouvoir expliquer la mentalité chez des êtres hypothétiques qui pourraient se situer en dehors de l'histoire évolutionniste.

Braddon-Mitchell et Jackson essaient de contourner les lacunes du fonctionnalisme analytique en affirmant que cette position peut en théorie établir ses règles, mais qu'en pratique, elle ne le peut pas, ou elle ne le peut pas encore. Comme le font les réductionnistes éliminativistes, les fonctionnalistes analytiques de la « deuxième heure » nous demandent de leur laisser la chance de tester leur théorie. Toutefois, malgré ma sympathie pour le fonctionnalisme en général, le pari éliminativiste semble dans ce cas-ci un pari plus intéressant car il a au moins le

mental states that are taken to define them. » (p.48).

mérite de proposer une éventuelle explication complète en pratique alors que le flou du sens commun semble être intrinsèque, ce qui rend toute définition stricte inaccessible selon les critères du fonctionnalisme analytique.

Comme le soulignent Braddon-Mitchell et Jackson, les fonctionnalistes analytiques de la première heure semblaient avoir un critère plus strict. Mais après analyse, il semble trop strict pour être en mesure de nous fournir assez de termes mentaux intéressants. Ainsi, voici comment David Lewis voyait le critère du sens commun.

Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses (...) Add (...) all the platitudes to the effect that one mental state falls under another - « tooth-ache is a kind of pain », and the like. Perhaps there are platitudes of other forms as well. Include only platitudes which are common knowledge among us -everyone knows them, everyone knows that everyone else knows them, and so on. (Lewis, 1972, p.256)

Même si à première vue ceci peut sembler un critère valable, il est en fait beaucoup trop fort²⁵. Existe-t-il vraiment plusieurs platitudes connues et acceptées de tous? Après tout, les Churchland ne sont-ils pas la preuve vivante (et ils ne sont pas les seuls) que les termes mentaux ne seront pas acceptés de la même manière par tout le monde. Aucun terme mental ne semble pouvoir faire l'unanimité demandée par Lewis. C'est pour cette raison que Braddon-Mitchell et Jackson préféraient un critère plus flexible. Mais dans les deux cas, le fonctionnalisme analytique semble ne pas être vraiment satisfaisant. Ou bien il est tellement strict qu'il ne pourra établir aucun objet d'étude, ou bien il est si généreux que ce qu'il explique devient trivial.

C'est pourquoi, différentes alternatives fonctionnalistes furent proposées. Braddon-Mitchell et Jackson offrent dans leur livre un bref survol des différentes avenues offertes au fonctionnaliste.

1.4 Fonctionnalisme empirique

Plutôt que de simplement dire que c'est la douleur qui me fait retirer ma main comme le dirait le fonctionnalisme du sens commun, le fonctionnalisme empirique offre deux avenues: Il affirme soit que la douleur correspond à un réseau neuronal

²⁵ Nous prendrons ici le commentaire de Lewis à la lettre. Peut-être que la caractérisation de ce genre de critère se veut ironique, toutefois, elle reste cohérente avec le reste de l'argument de Lewis alors

particulier qui lui joue un rôle fonctionnel, soit que la douleur correspond à un état fonctionnel particulier. Ainsi la douleur n'est pas un état fonctionnel de par sa nature, mais bien une manière de décrire les états fonctionnels.

La première avenue qui consiste à identifier un état mental à une explication neurophysiologique est selon Braddon-Mitchell et Jackson peu endossée aujourd'hui car elle rend impossible la réalisation multiple qui est un des fondements du fonctionnalisme. Toutefois la seconde voie reste intéressante: elle reste à un niveau d'abstraction assez grand pour permettre l'éventuelle attribution de la conscience à un Martien hypothétique. Cette voie a aussi l'avantage de maintenir la pertinence du discours neurophysiologique; elle fait ceci en traitant la description neurophysiologique comme une instanciation particulière d'un substrat pour le réseau fonctionnel désiré. Braddon-Mitchell et Jackson avancent, justement je crois, que l'on ne peut pas se débarrasser complètement de la description par le sens commun (ce que certains fonctionnalistes empiriques tentent de faire), car seul le sens commun peut faire sens de la source du questionnement fonctionnel. Ainsi, si j'éradique la douleur (telle qu'elle pourrait exister dans un savoir tacite du sens commun) du vocabulaire fonctionnaliste, on ne voit pas pourquoi on essaierait de comprendre comment fonctionne la douleur ou tout autre phénomène que le fonctionnalisme pourrait chercher à expliquer. Ainsi le rôle du sens commun dans les fonctionnalismes empiriques est soit absent soit limité à une espèce de mise en contexte du problème. La douleur ne fera pas partie du système fonctionnel en tant que tel, mais elle nous permettra de savoir ce dont on veut parler, ce que l'on veut expliquer, par une description fonctionnelle particulière. Sans ceci, le fonctionnalisme n'a plus de raison d'être.

Reste à savoir si le fonctionnalisme empirique peut fournir un critère plus intéressant que le fonctionnalisme analytique. L'intuition derrière le fonctionnalisme empirique est que la constitution physique du cerveau, ou d'autres systèmes, peut fournir certaines contraintes strictes (plus strictes que le simple critère d'un savoir tacite) pour déterminer quelles sont les fonctions pertinentes et leur rôle. Ces

contraintes strictes pourraient encore avoir une forme disjonctive (comme nous le verrons avec Shoemaker), mais elles seraient disjonctives à l'intérieur d'un domaine défini par la constitution physique, et non à l'intérieur d'un domaine dont la seule limite est l'imagination du sens commun.

Le problème avec cette voie de recherche est qu'en fixant les états fonctionnels dans la neurophysiologie on se coupe de la réalisation multiple extra-humaine. Ainsi, si on veut éviter le flou du fonctionnalisme analytique, en fixant les rôles fonctionnels non pas sur le sens commun, mais sur une certaine architecture physique, on pourra difficilement dire comment le Martien pourrait être conscient. C'est ce que les commentateurs appellent le chauvinisme de l'esprit (ce thème reviendra dans plusieurs articles que nous étudierons). Braddon-Mitchell et Jackson croient que tout fonctionnalisme empirique verse dans un chauvinisme de l'esprit en ne rendant possible la conscience que pour les humains. La seule manière d'éviter ce chauvinisme est de rendre très abstrait le lien entre le système fonctionnel et le substrat physique. Ce genre de fonctionnalisme empirique est appelé par Braddon-Mitchell et Jackson *machine functionalism* et est inspiré des *machine tables* qui correspondent aux programmes des machines de Turing. Le problème ici est que l'on perd la contrainte stricte désirée initialement par le fonctionnalisme empirique en général et qu'en se séparant largement des conditions de réalisations du système, le « programme » fonctionnaliste pourrait être instancié par à peu près n'importe quoi, ce qui rend toute explication fonctionnaliste triviale²⁶.

Sans aller jusqu'aux programmes des machines de Turing, le fonctionnalisme empirique pourrait soutenir que l'architecture précise ne compte pas, que seulement ce qu'elle fait importe. Le système fonctionnel ne correspondrait donc plus à l'architecture physique, mais bien aux actions causées par cette dernière. On ne voit plus très bien alors ce qui est particulier au fonctionnalisme empirique. En effet, sur quelle base pouvons-nous établir ce que fait cette architecture physique, si ce n'est le sens commun? Braddon-Mitchell et Jackson soulignent que le fonctionnalisme analytique semble être le plus prometteur. Le fonctionnalisme empirique est trop

²⁶ Dans notre examen de Sober (voir section 5 de ce mémoire), nous reviendrons au fonctionnalisme

chauvin ou trop libéral quant aux instanciations des systèmes fonctionnels. Si on essaie plutôt de trouver un juste milieu entre ces deux options, on réalise que le fonctionnalisme empirique se réduit au fonctionnalisme analytique.

Braddon-Mitchell et Jackson précisent que les différences entre les divers types de fonctionnalismes empiriques sont plus nuancées, mais il n'en reste pas moins qu'ils souffrent tous d'un certain degré de chauvinisme; or une des raisons d'être du fonctionnalisme est justement de s'éloigner de ce chauvinisme en rapatriant la réalisation multiple.

L'alternative s'offrant au fonctionnaliste se présente donc comme suit: il peut avancer que les états mentaux livrés par le sens commun *sont* des états fonctionnels (voie empruntée par le fonctionnalisme analytique ou de sens commun), ou que ces mêmes états ne sont en fait qu'une *description* d'états fonctionnels sous-jacents (voie empruntée par le fonctionnalisme empirique). Si le fonctionnaliste emprunte la deuxième option, il devra alors définir la nature exacte de ces états fonctionnels²⁷.

Ce simple examen des différentes manières de caractériser le fonctionnalisme et ses variantes montre la nécessité de préciser la thèse fonctionnaliste qui est visée par des arguments comme celui de l'inversion du spectre. Nous verrons tout au long de ce mémoire que cette précision est rarement apportée dans les arguments anti-fonctionnels, ce qui diminue leur efficacité. Toutefois, certains commentateurs semblent être conscients que cette nuance devrait être apportée. Comme nous le verrons dans la section suivante, Levine est un de ceux qui réalisent que le fonctionnalisme a plusieurs formes, mais il croit que l'argument de l'inversion du spectre attaque le fonctionnalisme à son fondement conceptuel, le matérialisme, ce qui rendrait caduque toute discussion quant aux différentes variantes fonctionnalistes.

machine et à ses lacunes.

²⁷ C'est à ce stade que se distingueront les diverses variantes empiriques.

2 Le gouffre explicatif

2.1 Argument épistémologique de Levine

2.1.1 Introduction

Après avoir vu comment le fonctionnalisme se présente sous ses différentes formes, il s'agit maintenant de voir si toutes ces formes peuvent être rejetées en bloc par un seul et même argument. Dans ce chapitre nous verrons la tentative de Levine en ce sens et la réplique d'un partisan du matérialisme en général, C.L. Hardin. La portée de l'argument de l'inversion du spectre est rarement appréciée à sa juste mesure par ceux qui l'utilisent. Comme Levine le montre dans son article « *Materialism and Qualia: The Explanatory Gap* » (1983), nous devons prendre conscience que tout argument reposant sur l'inversion du spectre constitue en fait, non seulement une attaque contre le réductionnisme, le béhaviorisme ou le fonctionnalisme, mais une attaque contre le matérialisme en général. Levine entend démontrer la limite explicative du matérialisme en général et il fait ceci en utilisant, entre autres, le cas de l'inversion du spectre. Les autres textes analysés dans ce mémoire ne visent pas aussi haut. Nous verrons que cette «humilité» réduit leur pertinence en tant qu'ils ne semblent pas rendre justice à la force de l'argument qu'ils utilisent ou qu'ils n'apportent pas les précisions nécessaires quant à leur cible réelle.

L'examen de ce texte nous sera utile car il nous permettra de comparer l'argument anti-matérialiste de Levine aux autres arguments anti-fonctionnalistes et de voir comment la portée de ceux-ci est beaucoup plus limitée que leurs auteurs ne le croient. Ensuite nous étudierons la réponse de Hardin qui, dans « *Qualia and Materialism: Closing the Explanatory Gap* » (1987), se servira d'un argument montrant l'impossibilité d'une inversion indétectable pour essayer de défaire l'argument anti-matérialiste de Levine. De plus, il montrera que les énoncés d'identité psychophysique ne sont pas plus contingents que des énoncés que Levine considère lui-même comme étant nécessaires, ce qui désamorcerait son scepticisme. Si cette «querelle» nous intéresse c'est qu'elle illustre pourquoi l'argument de l'inversion du spectre n'est pas simplement une attaque contre le fonctionnalisme

(ou contre quelques-unes de ces variations), mais une attaque contre toute thèse matérialiste. Elle peut donc être considérée non seulement comme un problème épistémologique mais aussi un problème métaphysique, bien que Levine ne prétende pas aller jusque là.

2.1.2 L'argument de Levine

Levine utilise l'argument kripkéen de « *Naming and Necessity* » contre le matérialisme. L'argument va comme suit: d'abord tout énoncé d'identité utilisant des désignateurs rigides, s'il est vrai, doit être vrai dans tous les mondes possibles. Or on peut concevoir qu'un énoncé psychophysique soit faux (dans un monde possible) donc ce ne peut pas être un énoncé d'identité vrai (Levine, 1983, p.354). Levine croit que l'argument est convaincant, mais que la conclusion métaphysique de Kripke quant à la défaite du matérialisme est trop forte. Ainsi, il faudrait plutôt utiliser l'argument d'un point de vue épistémologique, pour montrer que les identités psychophysiques, comme elles ne peuvent pas expliquer certains contre-exemples, laissent un gouffre explicatif (*explanatory gap*), ce qui contredit la supposée portée explicative du matérialisme. Qui plus est, comme le matérialisme n'est pas capable de rendre compte de certains cas possibles²⁸, il faut douter de tous les autres énoncés d'identité qu'il propose. Cette conclusion est épistémologique et non métaphysique.

En examinant deux énoncés d'identité physicaliste²⁹, (1) La douleur correspond à l'activation des fibres-C et (2) la chaleur correspond au mouvement des molécules, Levine dit que pour qu'ils soient vrais, ils doivent être *nécessairement* vrais. Or la nécessité de ces deux énoncés n'est pas intuitive. Au contraire, on peut facilement imaginer un monde possible où la chaleur correspond à autre chose qu'au mouvement des molécules et la douleur n'est pas associée à des fibres ou réseaux

²⁸ Il est important de souligner qu'il y aurait une distinction à faire entre les cas possibles et les cas concevables. Cette distinction semble tourner autour d'une distinction entre métaphysique (cas possibles) et épistémologique (cas concevables). Toutefois je n'approfondirai pas cette question pour deux raisons. D'une part, cette question constitue un débat complexe en soi, mais surtout, les auteurs s'étant intéressés à la question de l'inversion du spectre ont la plupart du temps utilisé les deux termes indifféremment. Cette lacune est peut-être un signe que l'argumentation est déficiente dès le départ, mais j'y vois plutôt un espoir que les conclusions épistémologiques qu'ils tireront ont aussi des conséquences métaphysiques.

²⁹ Pour tout son argument, Levine ne se concentre que sur les identités type-type et non token-token.

neuraux. Toutefois, comme le souligne Kripke, la contingence de (2) peut être abolie si on réalise que l'énoncé que l'on juge en réalité contingent n'est pas (2), mais un (2)' que l'on pourrait définir comme étant « le phénomène que l'on sent comme étant chaud ou froid correspond au mouvement des molécules ». Ainsi, ce ne sont pas les énoncés d'identité physiques que nous trouvons en réalité contingents, mais les énoncés d'identité psychophysiques³⁰. Toutefois (1) n'est pas un énoncé « objectif » comme (2) pourrait sembler l'être, car on ne peut faire de distinction entre la douleur « subjective » que nous sentons et la douleur « objective »; il n'y a que la douleur que nous sentons, alors qu'il y a la chaleur en tant que phénomène physique et la chaleur que nous sentons (et ces deux phénomènes sont distincts). Comme on peut imaginer un monde possible où la douleur n'est pas associée à l'activation de fibres-C, (1) n'est pas vrai dans tous les mondes possibles, et donc n'est pas vrai au sens kripkéen. Levine croit que tous les énoncés d'identité psychophysique sont condamnés au même sort.

Si l'identité psychophysique échoue, le matérialiste peut espérer s'en sortir en favorisant une explication fonctionnaliste plutôt que réductionniste. Ainsi nous aurions (3) « avoir de la douleur » qui correspond à être dans l'état F. Levine, s'appuyant sur l'argument de Block³¹, pense que ce genre d'énoncé ne sauve pas le matérialisme car le même reproche de contingence s'applique à ce genre d'énoncés (on peut imaginer un système n'ayant pas cet état F ou l'ayant mais n'ayant pas de douleur). Ainsi, la reformulation en (3) n'implique pas plus la nécessité que ne l'est capable (2)'.

Levine cherche aussi à étoffer les différences entre les énoncés d'identité psychophysique (comme (1) et (3)) et les autres types d'énoncés réductionnistes physiques (comme (2)). Même si la contingence des énoncés psychophysiques constitue le fer de lance d'une argumentation de style kripkéen, Levine croit que ceux-ci comportent une autre faille qui fait douter de leur validité. En effet, alors qu'un énoncé comme (2) (la chaleur correspond au mouvement des molécules)

³⁰ Ce qui fait que (2)' est un énoncé d'identité psychophysique est qu'on le définit en termes de *notre* expérience sensorielle et non en tant que phénomène objectif.

³¹ Voir l'argument de la Chine comme émulation fonctionnelle de l'esprit humain dans « *Troubles*

semble être capable d'épuiser ce qu'il y a d'important à dire au sujet de la chaleur, les énoncés (1) et (3) semblent tous deux oublier des caractéristiques fondamentales des phénomènes qu'ils cherchent à expliquer. C'est pour cette raison que Levine dira que les énoncés d'identité psychophysique comportent un gouffre explicatif (*explanatory gap*) entre ce qu'ils veulent expliquer et ce qu'ils expliquent en réalité.

Levine n'affirme pas que les explications fonctionnalistes du type « la douleur correspond à l'activation de fibres-C » sont fausses, mais bien qu'elles sont incomplètes car elles n'arrivent pas à rendre compte des qualia associés³². C'est ici que l'argument de l'inversion du spectre entre en jeu. Il définit le problème de la manière usuelle: l'agent normal est dans l'état fonctionnel V lorsqu'il voit un concombre, et il a le qualia du vert. Mais on ne semble pas être en mesure de rejeter la possibilité d'un agent qui serait dans l'état fonctionnel V lorsqu'il verrait un concombre, mais qui aurait le qualia du rouge. Cette contingence démolirait le fonctionnalisme. On pourrait alors utiliser un argument cherchant à fonder le fonctionnalisme dans le physicalisme (voir les arguments de Shoemaker plus loin dans ce mémoire), mais resterait toujours le problème de savoir pourquoi ou comment se fait-il que nous ayons ces expériences comme nous les avons plutôt que d'être de simples zombies sans expérience phénoménale. Levine ne pense pas qu'une explication physique (ou fonctionnaliste) soit en mesure de fournir une explication satisfaisante des propriétés phénoménales et l'inversion du spectre en serait un exemple flagrant. Nous sommes possiblement incapables de découvrir « fonctionnellement » si un agent voit vert ou rouge, alors on peut imaginer que de tels qualia ne soient pas associés aux états fonctionnels étudiés. Ceci veut dire que le fonctionnalisme ne peut expliquer en entier la vie mentale des agents, ce qui le condamne autant que son prédécesseur, le béhaviorisme (ou toute autre thèse matérialiste).

with Functionalism » et dans la section 3.1 de ce mémoire.

³² « *However, there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is why should pain feel the way it does!* » (Levine, 1983, p.357).

2.1.3 Conclusion

Une approche fonctionnaliste analytique ne semble pas être en mesure de prouver la nécessité des identités psychophysiques à travers tous les mondes possibles, d'autant plus qu'en se détachant des conditions de réalisation particulières retrouvées dans le cerveau humain, il semble difficile de pouvoir fonder des systèmes fonctionnels qui ne sont pas sujets à cette critique de contingence. Certains fonctionnalistes ont donc tenté d'ancrer leur fonctionnalisme dans des réalisations multiples physiques. Ainsi, la douleur correspondrait à UN état fonctionnel (ceci n'est bien sûr qu'une caricature grossière des mécanismes impliqués) qui puisse être réalisé physiquement de diverses manières et ce sont ces réalisations particulières qui font que ce qu'on pourrait appeler des « qualia » diffèrent d'une personne à une autre. C'est le genre d'argument proposé par Shoemaker dans « *Inverted Spectrum* » (1982, l'argument sera examiné en détails dans la section 3.1 de ce mémoire). Toutefois, le problème kripkéen demeure car l'association de l'expérience de douleur à UN état fonctionnel associé à DES réalisations physiques multiples reste contingente comme le montre entre autre le cas de l'inversion du spectre. La seule manière de pouvoir intégrer les réalisations physiques à cette description fonctionnaliste est de définir les qualia de manière disjonctive. Le problème est que cette disjonction est définie en termes de nos états et est donc relativement limitée. Or, si un extraterrestre nous dit souffrir, mais qu'il n'a pas de fibres-C, serions-nous obligés de dire qu'il ne souffre pas vraiment? Si on dit qu'un agent souffre, qu'il est dans l'état fonctionnel S réalisé par l'état physique A OU B, pourquoi limiter les réalisations physiques à A et B? Le fonctionnalisme ne fournit pas un critère d'évaluation de similarité ou de différence entre les états (quel degré de similarité entre deux états doit être atteint pour que l'on puisse affirmer que deux agents sont dans le même état?).

Ainsi, la disjonction introduite par les fonctionnalistes « physicalistes »³³, loin de sauver le fonctionnalisme, introduirait une disjonction qui rendrait la classe des réalisations physiques associées à un état, soit trop stricte a priori (la définition

³³ le fonctionnalisme physicaliste fait partie des fonctionnalismes empiriques d'après la nomenclature de Braddon-Mitchell et Jackson.

disjonctive serait trop restreignante et empêcherait plusieurs états *analogues* d'être inclus dans cette description fonctionnaliste particulière), ou trop généreuse (admettant des états qui ne sont pas vraiment analogues) et donc triviale. Ce problème sera soulevé à plusieurs reprises tout au long de ce mémoire.

Pour dissiper la contingence, il faudrait que les qualia soient définis à l'intérieur des propriétés physiques qui cherchent à les expliquer ce qui reviendrait à éliminer les qualia (position éliminativiste à la Churchland). Levine affirme que c'est la seule manière pour le matérialiste de sauver sa position car il ne pourra pas rendre compte des qualia ET des états physiques en même temps. Toutefois, Levine ne pense pas que cela soit une solution satisfaisante car elle va à l'encontre du sens commun sur la question de la vie mentale des agents³⁴.

2.2 Critique de Hardin

2.2.1 Introduction

Bien entendu, les matérialistes ne se laisseront pas faire devant une attaque si fondamentale de leur position. C.L. Hardin dans « *Qualia and Materialism: Closing the Explanatory Gap* »(1987) essaie de répondre à Levine. Hardin s'intéresse surtout au problème de la couleur qu'il ne juge pas semblable à l'exemple de la chaleur perçue par un agent (le (2)' kripkéen). Ainsi, Hardin ne semble pas croire que l'inversion du spectre puisse être traitée comme une simple instanciation du problème général de l'inversion des qualia, mais plutôt comme un problème autonome qui aura sa résolution particulière. Ainsi, sa réponse ne cherche pas à détruire les qualia en général, mais bien à montrer que le matérialisme ne connaîtra pas la déconfiture face à l'argument de l'inversion du spectre. Toutefois, et on sent cet espoir aussi dans ses autres textes, Hardin semble croire que ce genre de solution pourrait être appliqué à d'autres problèmes liés aux qualia³⁵. Provisoirement, on peut dire qu'il cherche au moins à montrer que l'on ne peut pas traiter les énoncés

³⁴ Bien sûr, les positions éliminativistes habituelles ne voient pas là une lacune car le sens commun s'est, par le passé, trop souvent trompé à propos des phénomènes naturels (mentaux ou autres). Il ne faudrait donc pas utiliser le sens commun comme critère de plausibilité réelle d'une théorie.

³⁵ Il y a donc une tension chez Hardin entre le désir de distinguer le problème de l'aperception chromatique d'autres expériences phénoménales et le désir d'affaiblir les arguments anti-matérialistes

d'identité psychophysique comme faisant partie d'une classe homogène, qui peut être détruite ou sauvée à l'aide d'un seul genre d'argument.

2.2.2 Couleur et physicalisme

Hardin base son argument sur une certaine conception de la couleur. Il commence par préciser que la couleur n'est pas simplement fonction de propriétés physiques objectives des objets. Ainsi, ce n'est pas seulement parce qu'un objet reflète une certaine longueur d'onde particulière que nous percevons du rouge. Hardin identifie plusieurs autres facteurs qui influencent la perception chromatique de l'agent et plusieurs de ces facteurs concernent l'appareil sensoriel de l'agent et non pas seulement des propriétés externes à l'agent³⁶. Hardin croit que c'est la seule manière « honnête » de considérer la couleur, car faire de la couleur simplement une propriété externe, pour éviter le problème des qualia (en essayant d'objectiviser la perception des couleurs), ne peut pas rendre compte de notre expérience. Hardin cherchera à camper sa thèse contre la position « dualiste » (que l'on identifiera aux pro-qualia³⁷) et, plus surprenamment, contre le fonctionnalisme³⁸. Il ne croit pas que sa position puisse être intégrée par le fonctionnalisme, puisque cette théorie rejeterait une explication de la perception fondée exclusivement sur les conditions de réalisation physique de la perception. Ainsi il pense que les fonctionnalistes rejetteraient son genre de caractérisation de la perception chromatique comme étant trop chauvin ou « *parochial* »³⁹ car elle ne pourrait pas rendre compte de la vie

en général grâce à son argument concernant l'inversion du spectre.

³⁶ Hardin se servira de l'énumération présentée par Hurvich (1981). « *What we should keep in mind is that the color that will be seen by a particular person in a particular situation will be a function of a whole host of variables over and above the spectral characteristics of the stimulus. To consider a simple, well controlled circumstance, suppose that a « normal » subject is stimulated by a flash of light of a given wavelength composition. The color that she will see will depend upon (1) the state of adaptation of the eye, (2) the absorptive properties of the ocular media, (3) the absorptive properties of the receptors, (4) the portion of the retina on which the flash falls, (5) the efficiencies of the postreceptor neural processes, (6) the intensity of the flash, (7) the angular size of the flash, (8) the duration of the flash, (9) the background against which the flash is seen.* » (Hardin, 1987, p.283)

³⁷ « *The dualists argue that the intrinsic natures of the phenomenal features of mental life make them in principle irreducible to their biological substratum.* » (Hardin, p.284)

³⁸ « *The functionalists, for their part, are inclined to hold that hue or tone perceptions are to be characterized primarily by how the perceiving organism discriminates and classifies particular sorts of stimuli in its environment. The particular realizations of these capabilities play a secondary role in understanding the related mental states.* » (Hardin, p.284)

³⁹ Le mot « *parochial* » est très souvent utilisé dans la littérature fonctionnaliste et pourrait être traduit

mentale des hypothétiques martiens ou d'un robot très développé, deux exemples prisés par le fonctionnalisme « pur ».

Même s'il est clair que la position physicaliste de Hardin ne peut être accommodée dans le fonctionnalisme « pur » (fonctionnalisme analytique ou de sens commun), cette incompatibilité semble davantage relever d'un combat contre un « *strawman* » que contre un ennemi réel. En effet, plusieurs positions fonctionnalistes se sont distancées de l'abstraction totale retrouvée dans les premières explications fonctionnelles, préférant emprunter une voie un peu plus réductionniste. Dans ses écrits plus récents, Hardin a concédé que sa position pourrait être comprise comme une position fonctionnaliste, même s'il ne voit pas sa position comme l'étant en réalité (voir Hardin, 1991, p.45, et la section 4.2 de ce mémoire où je reviens sur la question).

L'argument de Hardin contre Levine sera donc une description physicaliste « pure » de la perception chromatique, cherchant ainsi à prouver que le matérialisme survit à l'argument de l'inversion du spectre. La croyance que la compréhension de la perception chromatique ne puisse pas être accomplie par les sciences relève, selon Hardin, de notre ignorance relative sur la question et il nous exhorte donc à le suivre dans son explication physicaliste de ces mécanismes pour voir si le rejet *a priori* effectué par Levine est valide.

2.2.3 Spéculation et « *opponent process theory* »

Hardin croit que l'argument de Levine est fort analogue à l'argument du moulin de Leibniz (voir « *Monadologie* », section 17). Ainsi même si le cerveau était gros comme un moulin et qu'on pouvait ausculter son intérieur et voir ses « rouages », on ne pourrait pas inférer de ces observations que cet immense cerveau a des expériences phénoménales. Hardin croit que le moulin de Leibniz et l'inversion du spectre de Levine pèchent par leur manque de profondeur. Les deux arguments décrivent une science où les structures ne sont décrites que de manière relativement macroscopique et grossière comme si les structures de perceptions n'étaient décrites

par « avoir un esprit de clocher » trop centré sur les problèmes de sa paroisse au détriment d'une vue d'ensemble. On dit qu'une position est chauvine ou « *parochial* » si elle ne peut, ou ne cherche, qu'à

qu'en examinant des neurones ou un agencement simple de ceux-ci. Hardin défend l'idée qu'il existe une théorie psychophysique particulière qui est capable de rendre compte de notre expérience phénoménale de la couleur: *l'opponent process theory* (par la suite abrégé OPT). Pour les besoins de l'argument et pour nos besoins subséquents, voici une brève description de cette théorie (il est important de souligner que cette théorie n'a pas été établie par Hardin, mais bien par des scientifiques de la couleur; on peut trouver la caractérisation originale de Hardin, p.287 inspirée de la description plus complète offerte par Hurvich, 1981).

Le rouge et le vert seraient codés sur le même canal chromatique (appelé le *r-g channel*) alors que le bleu et le jaune seraient codés sur un autre canal (appelé le *y-b channel*) et le noir et le blanc sur autre canal. Chaque canal a un état d'activité de base (lorsqu'il n'est pas stimulé) qui code pour un gris foncé. Si on excite un de ces canaux on obtient une couleur, alors que si on inhibe un canal on obtient sa couleur complémentaire. Comme on ne peut pas exciter un canal et l'inhiber en même temps, on ne pourra pas avoir du vert et du rouge simultanément. Une couleur codée sur un seul canal est dite unique, alors qu'une couleur qui est le résultat de l'activité de deux canaux est dite binaire⁴⁰. Ce genre de théorie empêche donc la possibilité d'inversion indétectable de *certaines* couleurs. Ainsi si on tentait d'inverser, par exemple, le rouge et l'orange, il faudrait rejeter OPT. En effet, l'orange deviendrait une couleur unique et le rouge une couleur binaire, ce qui rendrait les autres couleurs impossibles (ou du moins très difficiles) à coder. Ainsi, en admettant l'inversion de l'orange et du rouge, comment peut-on coder le mauve? On ne pourrait pas coder le mauve à partir de l'inhibition du canal *y-b* ET à partir de la couleur *maintenant* binaire qu'est le rouge. Bien sûr, on pourrait encore soutenir que, dans un monde possible particulier, les humains perçoivent la couleur différemment et que ces différences rendent possibles une inversion du rouge et de l'orange (ou d'autres types d'inversion). Ici Hardin offre une critique des arguments *a priori* en général. Si on veut toujours affirmer la contingence du genre d'explication proposée par OPT, il

expliquer la mentalité humaine.

⁴⁰ « For instance, purple is signaled when the *r-g* channel codes for red and the *y-b* channel codes for blue. If, on the other hand, one sort of channel codes for a hue while the other sort of channel fires at

faut aussi affirmer la contingence d'affirmations comme la chaleur correspond au mouvement des molécules (car on peut imaginer un monde où la chaleur ne correspond pas au mouvement des molécules, mais à un autre phénomène). Or le but de Levine n'était pas de rejeter tout réductionnisme, mais bien de rejeter les énoncés d'identité psychophysique. Mais si Levine veut affirmer la contingence d'explication comme OPT, il devra aussi rejeter des énoncés qu'il avait auparavant acceptés comme étant nécessaires (comme son énoncé 2). Ainsi, la voie que peut emprunter Levine pour permettre l'inversion de n'importe quelles couleurs, ouvre l'*explanatory gap* à tous les savoirs scientifiques que l'on pourrait juger non-contingents. Le dualisme de Levine semble mener à un scepticisme dépassant le sens commun que celui-ci prétend défendre. Le dualiste serait condamné non seulement à douter des énoncés d'identité psychophysique mais à une pléthore d'autres énoncés qu'il jugerait *nécessairement* vrais auparavant.

Toutefois, Hardin est bien conscient des limites internes de son argument: rien dans l'OPT n'empêche l'inversion de couleurs uniques entre elles. Ainsi, le rouge, plutôt que d'être le résultat de la stimulation du canal *r-g*, pourrait être le résultat de son inhibition, et le vert vice versa. L'OPT empêche la possibilité d'inversion de *certaines* couleurs entre elles, mais pas de toutes; les deux opposés d'un même canal pourraient être inversés sans détruire tout le codage chromatique. Toutefois ces inversions seront détectables. Alors que son argument précédent empêchait la possibilité de l'inversion sur la base du codage des couleurs, Hardin utilisera ici des propriétés associées aux couleurs pour montrer que l'inversion n'est pas impossible, mais qu'elle est détectable de manière comportementale, ce qui sauverait le matérialisme (et, par accident, le fonctionnalisme).

Hardin appuie son argument sur le fait que les couleurs semblent être associées à des impressions de chaleur ou de profondeur (ou à d'autres propriétés). Ainsi, le rouge semble être une couleur « chaude » alors que le bleu semble être une couleur « froide ». Le fait que la couleur ait des propriétés intermodales permettrait de déceler l'inversion chez un individu. Si on demande à un agent de dire ce

its base rate, the output is the code for a non-composite, or unique hue. » (Hardin, p.287)

qu'évoque le rouge d'un panneau d'arrêt et qu'il dit le froid, nous aurons un indice d'inversion, car les agents « normaux » associeront plutôt la chaleur à cette couleur. Hardin veut montrer que même si l'inversion est possible, elle ne serait pas indétectable à cause des propriétés *non-chromatiques* qui seraient intrinsèquement liées aux perceptions de couleurs. La chaleur n'est pas la seule propriété associée. Mais que ce soit pour n'importe quelle propriété associée, on observe qu'il y a une polarité entre la propriété associée à un des pôles d'un des canaux et la propriété associée au pôle opposé du même canal. Le rouge, en tant que pôle « positif » (car il est le résultat d'une stimulation), est considéré comme étant chaud alors que le vert, qui est le pôle « négatif » (car il est le résultat d'une inhibition), est considéré plus froid. La même polarité apparaît sur les autres canaux. Le jaune qui est « positif » est chaud alors que le bleu « négatif » est froid. L'inversion serait détectable car elle donnerait lieu à des jugements étranges quant aux propriétés associées.

Ce genre de solution est très séduisant car il offre un moyen de résolution matérialiste tout en admettant la possibilité de l'inversion. C'est pour cette raison que plusieurs philosophes qui essaient de naturaliser l'esprit ont aimé l'idée de Hardin. Toutefois, on ne sent pas toujours la prudence qui devrait accompagner une telle position. Flanagan, dans « *Consciousness Reconsidered* » (1992), se base sur l'argument de Hardin pour sauver le matérialisme de l'inversion du spectre. Le but de Flanagan est clair: préserver le quale et les expériences de pensée qu'il amène, tout en fournissant un arrière-plan matérialiste pour expliquer le tout. Cette avenue comporte une tension interne implacable. Ainsi, Flanagan souligne que des moyens analogues à ceux de Hardin dévoileront les cas d'inversion; et le matérialisme est donc sain et sauf. Toutefois, il admet que le quale gardera une certaine part d'« inscrutabilité »⁴¹, ce que n'admet pas la thèse radicale de Hardin. Le réductionnisme de Hardin a au moins le mérite de ne pas laisser place à cette inscrutabilité. Bien que cette option ne soit pas très intuitive, elle a au moins le mérite d'être cohérente alors qu'il n'est pas clair que la thèse de Flanagan puisse

⁴¹ « *The idea that a neuroscientist armed with a good understanding of the neurophysiology of color vision and PET or MRI scanners could detect such [inversions] is perfectly plausible.(...)We will never be in a position to prove that two individuals see red in exactly the same way.* » (Flanagan,

supporter cette inscrutabilité partielle tout en affirmant vouloir complètement naturaliser l'esprit.

Le plus grand problème de l'argument de Hardin est qu'il semble être, au mieux, optimiste. Les faits portant sur les propriétés associées sont discutables⁴². En fait, Hardin nous demande de suivre son intuition: « *The pervasiveness of these phenomena suggests that their roots may be embedded deeply within our physical makeups.* » (Hardin, p.289). Le langage employé par Hardin est prudent, mais les *pervasiveness, suggests, may be*, semblent bien faibles pour fonder ces propriétés associées dans la structure physique du cerveau. Bien sûr Hardin ne s'arrête pas là et fournit certaines pistes permettant de postuler des cellules comparatrices intermodales qui permettraient d'associer la polarité des canaux chromatiques à d'autres sensations. Je n'ai pas toutes les connaissances nécessaires pour juger de ses affirmations relatives à la constitution et l'évolution du système nerveux, mais certains commentaires restent possibles: l'affirmation de Hardin à l'effet que les stimuli chromatiques ont une grande influence sur le comportement des vertébrés et des invertébrés ne me semble pas problématique en soi (bien que des preuves empiriques à l'appui auraient dû être offertes par Hardin). Toutefois, l'idée que le rouge est *nécessairement* associé à la chaleur me semble injustifiée. Car comment peut-on prouver que cette association n'est pas en réalité culturelle, et donc contingente⁴³?

2.2.4 Cellule comparatrice et inversion du spectre

Examinons la manière dont Hardin se sert de cette cellule comparatrice pour sauver le matérialisme. Si on accorde le bénéfice du doute à Hardin et si une telle cellule existe (ainsi que les phénomènes intermodaux qu'il postule), un cas d'inversion ressemblerait à une des trois options suivantes.

p.72)

⁴² Hardin soulève lui-même le problème: « *There has been considerable difficulty in obtaining unambiguous results here, largely because of methodologically flawed data.* » (Hardin, p.289)

⁴³ Levine offrira dans son article « *Cool Red* » (1991) un argument différent contre la thèse de Hardin: ainsi, même si le rouge et la chaleur étaient associés physiquement dans notre cerveau, on pourrait imaginer un agent chez qui cette association est inversée aussi (en fait, il inversera les cellules comparatrices intermodales). J'examine cet argument et la réplique de Hardin dans la section 4 de ce mémoire.

(1) L'agent supposément inversé dirait qu'une tomate mûre (rouge pour l'observateur « normal ») est « froide ». Étant donné les associations hypothétiques standards entre certaines couleurs et certaines réactions émotives, on pourrait en déduire que l'agent ne perçoit pas la même chose que nous et, donc, l'inversion ne serait pas indétectable comportementalement. Le matérialisme en général serait sauf car il y a une asymétrie comportementale entre l'agent normal et l'agent inversé.

(2A) L'agent supposément inversé dirait que la tomate mûre est « chaude » comme le dirait l'agent normal. Cette option nous mène à la conclusion peu attirante que les qualia sont des épiphénomènes car ils ne modifient pas les comportements. Les qualia ne « servent » donc à rien et sont détachés de ses croyances associées à ses perceptions.

(2B) L'agent inversé aurait des perceptions inversées ET une cellule comparatrice inversée. Ainsi il aurait un quale du vert en voyant une tomate mûre; mais sa cellule comparatrice étant inversée elle aussi, le vert serait associé au chaud et l'agent dirait donc que la tomate est « chaude ». Cette double inversion ferait que le comportement de l'agent ne pourrait pas révéler l'expérience phénoménale de l'agent.

Ce que Hardin veut montrer est que, si l'inversion est possible, 1- ou bien elle est nécessairement détectable, 2- ou bien elle mène à des conséquences contre-intuitives tel l'épiphénoménisme de notre expérience perceptuelle, 3- ou elle conduit à la conclusion étrange de 2B. 2B est étrange car il faut se demander ce qu'il reste d'intrinsèque aux concepts chromatiques. Ainsi, si on peut inverser les perceptions, inverser les associations intermodales, que pourraient bien signifier « être vert » ou « paraître vert » autant pour l'agent normal que pour l'agent inversé? 2A et 2B veulent illustrer à quel point le dualisme, loin d'être la position la plus intuitive, est en fait la position la plus difficile à accepter, car cette dernière, en se servant de l'inversion du spectre, montre que l'expérience phénoménale ne modifie pas nécessairement nos croyances et/ou que nos croyances perceptuelles ne déterminent pas nos comportements. Hardin ne nie pas la possibilité de l'inversion du spectre, mais bien la pertinence du dualisme qui encourage cette possibilité.

2.2.5 Conclusion

Comme l'inversion reste possible, il y aurait en effet un gouffre explicatif, mais il ne serait pas différent du gouffre explicatif qui existe entre n'importe quelle théorie et les faits qu'elle essaie d'expliquer. Levine ne peut donc pas distinguer la pertinence d'une explication de la température par le mouvement des molécules, de la pertinence d'une explication physicaliste de la perception. La conclusion de Hardin est claire: la résolution de ces problèmes est empirique, et on ne peut pas distinguer ce type d'explication des autres explications de la science.

Dans la dernière partie de son article Hardin examine le genre de réduction physicaliste qui s'offre à nous dans nos tentatives d'explication de nos perceptions. Étrangement, alors qu'il tentait auparavant de montrer pourquoi les perceptions n'étaient pas si différentes des autres phénomènes physiques que la science essaie d'expliquer, il veut maintenant souligner la spécificité de nos perceptions pour montrer que la réduction, bien que souhaitable, sera difficile. On ne peut pas observer nos perceptions; nous avons nos perceptions. Cette difficulté semble pouvoir être contournée si le réductionniste essaie plutôt d'identifier les conditions de vérité de ces perceptions en fonction de l'activation de certains réseaux neuronaux. « *Visual phenomenal talk would thus become a façon de parler for complex physiological transactions. Such a reduction of colors really amounts to their ultimate elimination.* »(Hardin, p.294). Bien sûr, Hardin est conscient que ceci ne sera pas suffisant. Il souligne que cette réduction devra aussi expliquer notre incapacité de retour réflexif sur ces perceptions (pourquoi ne pouvons-nous pas les observer directement comme tout autre objet d'étude). Bien que je sois en accord avec cette visée de recherche, je ne suis pas convaincu qu'elle puisse être accommodée dans un cadre purement réductionniste. Ainsi pouvons-nous vraiment nous dire éliminativistes si une part de la dite explication montre pourquoi nous sommes incapables d'expliquer totalement la nature de nos perceptions? Le fonctionnalisme, en essayant de se distancer de l'éliminativisme, me semble plus en mesure de fournir un cadre permettant ce genre d'explication. Ainsi un « gouffre explicatif » pourrait jouer un rôle causal dans la vie mentale des agents sans que l'on ait à éliminer ce gouffre pour le remplacer à tort, ou du moins d'une manière

décevante et contre-intuitive, par une supposée identité physique.

Hardin n'est pas *a priori* contre ce genre d'explication, mais il ne croit pas que son genre de résolution soit accessible au fonctionnalisme « pur ». La porte reste toutefois ouverte aux autres genres de fonctionnalismes empiriques⁴⁴. L'argument de Hardin est clairement empirique. Seul un fonctionnalisme qui tient compte de la réalisation empirique des systèmes qu'il propose pourra possiblement accommoder l'explication de Hardin. Peut-être qu'un fonctionnalisme téléologique serait le meilleur compromis entre le réductionniste éliminativiste de Hardin et le fonctionnalisme « pur ». En adoptant une telle position, on aurait accès au genre d'explication empirique de Hardin en tant qu'elle cherche à comprendre le but ou la raison d'être des constituantes du système étudié, tout en permettant une réalisation parallèle de ces systèmes dans des organismes non humains, par exemple. Les divergences entre le fonctionnalisme et la position de Hardin se rétréciront dans ses écrits postérieurs. Une conclusion provisoire à tirer de cet examen est que Levine et Hardin croient tous les deux que le fonctionnalisme « pur » (dit analytique ou de sens commun) n'est pas en mesure d'expliquer adéquatement les perceptions, et donc la vie mentale en général.

⁴⁴ « *How does this reductive materialism stand vis-à-vis functionalism? This is difficult to answer because of the wide range of positions which have been accounted functionalist, and because of the general floppiness of such key terms as « input », « output », and « mental state ». But insofar as functionalists prefer to generalize across broad classes of organisms and even machines, the differences are evident.* » (Hardin, p.298)

3 Inversion intersubjective et intrasubjective

3.1 Argument de Shoemaker en faveur de la possibilité de l'inversion

3.1.1 Introduction

L'argument de Hardin était clairement empirique. Toutefois, certains commentateurs ont essayé de sauver le fonctionnalisme de manière purement *a priori*. Shoemaker est un de ceux-là. La caractérisation de la thèse de Shoemaker dans mon mémoire est problématique. D'une part, Shoemaker affirme être fonctionnaliste et il fournit ce qu'il croit être des explications fonctionnalistes de différents mécanismes perceptuels ou phénoménaux. D'autre part, il concède à plusieurs reprises que les qualia ne peuvent probablement pas être décrits fonctionnellement: son affirmation que l'inversion du spectre est une possibilité logique est un pas dans ce sens. Par contre, Shoemaker ne croit pas que la cohérence de l'argument de l'inversion du spectre doive être un obstacle au fonctionnalisme. Il croit plutôt que, même si les qualia en eux-mêmes ne sont pas « fonctionnalisables », la relation de similarité ou de différence entre qualia et la classe des qualia est « fonctionnalisable », ce qui sauverait le fonctionnalisme contre une attaque basée sur l'inversion du spectre. Sydney Shoemaker défend une position mitoyenne qui essaie de rendre justice à certaines intuitions quant aux qualia tout en tentant de maintenir une forme affaiblie du fonctionnalisme. Il s'agira de voir si sa définition du quale correspond vraiment à une définition acceptable pour ses tenants, et de comprendre quelle est la forme de l'explication fonctionnaliste apportée.

3.1.2 L'argument métaphysique de Shoemaker

Shoemaker croit que le problème de l'inversion du spectre est en fait un double problème. Le front épistémologique, traité brièvement à la fin de « *Inverted Spectrum* » (Shoemaker, 1982), s'intéresse à la possibilité d'acquérir un savoir vrai concernant la vie mentale des autres agent. Ainsi, la question est de savoir si je suis en mesure de dire que mon voisin, lorsqu'il voit une tomate mûre, voit vraiment « rouge ». Shoemaker croit que ce problème se réglera ultimement de manière empirique (en fait, il ne croit pas qu'en réalité des cas d'inversion du spectre soient

possibles dans *notre* monde). Shoemaker s'intéresse surtout ici au second front du problème de l'inversion du spectre, soit l'aspect métaphysique. L'aspect métaphysique ne peut pas être réglé de manière empirique, car il ne s'intéresse pas à notre constitution particulière, mais bien à la définition conceptuelle des qualia, et tente de montrer comment cette définition affecte les portées explicatives du fonctionnalisme. Ainsi, même si on prouvait épistémologiquement que l'inversion n'est pas possible chez les humains, l'argument métaphysique resterait entier: si on peut imaginer deux créatures (humaines, martiennes ou autres) fonctionnellement isomorphes, mais ayant des qualia distincts, comment peut-on soutenir la valeur explicative du fonctionnalisme? Alors que Shoemaker ne croit pas que l'inversion du spectre existe dans les faits (problème épistémologique), il croit qu'on ne peut pas nier la possibilité logique de l'inversion du spectre (problème métaphysique).

L'argument métaphysique constitue la formulation la plus usuelle des critiques anti-fonctionnalistes. Selon les anti-fonctionnalistes la simple possibilité logique de l'inversion du spectre met en danger le fonctionnalisme. Contrairement à plusieurs fonctionnalistes, Shoemaker ne cherchera pas à montrer l'impossibilité de l'inversion du spectre, mais de voir si cette possibilité détruit vraiment le fonctionnalisme. Il est important de noter que contrairement à plusieurs commentateurs qui placent l'inversion du spectre visuel dans la même classe d'arguments que celle de l'absence de qualia, Shoemaker refuse la possibilité de l'absence de qualia, tout en admettant la possibilité d'une inversion du spectre (à ma connaissance, il est le seul à le faire).

Un argument utilisant l'absence de qualia a été présenté par Block. Il propose que, si la population de la Chine se divisait les tâches d'un cerveau (chaque Chinois agissant comme une espèce de neurone), cette émulation, bien que capable d'imiter toutes les actions du cerveau, ne serait pas consciente (l'argument basé sur la possibilité des zombies, des êtres totalement identiques à nous, mais dépourvus de conscience, va dans le même sens). Block affirmait que le cerveau et la Chine seraient alors dans des configurations fonctionnelles similaires à celle de l'homme,

bien que, dans le cas de la Chine, il n'y aurait pas de conscience. Cet argument vise à montrer les limites explicatives du fonctionnalisme.

Pour sa part, Shoemaker croit pouvoir éliminer la possibilité conceptuelle de l'absence de qualia. Il affirme que tout qualia donne lieu à une « croyance qualitative », une croyance par rapport aux qualia. Ainsi, lorsque j'ai mal (qualie de douleur) ma conscience introspective de cet état donne lieu à une croyance quant à mes qualia. Même s'il n'est pas clair que les qualia puissent jouer un rôle fonctionnel, il est évident, pour Shoemaker, que nos croyances relatives à nos états qualitatifs jouent un rôle causal dans notre comportement. Ainsi, dire qu'un système pourrait être fonctionnellement identique à un autre sans reproduire les qualia de ce dernier, c'est dire que nos croyances à propos de nos qualia ne jouent aucun rôle fonctionnel, ce qui va à l'encontre le sens commun (notre conscience par rapport à nos qualia semblent fréquemment modifier nos comportements).

Dans les cas où on croit avoir un exemple d'absence de qualia, il n'y a donc pas en fait d'isomorphie fonctionnelle entre les deux systèmes, et le fonctionnalisme reste sain et sauf. Un état fonctionnel sans caractère qualitatif ne peut pas être comparé à un état fonctionnel accompagné de qualia, car ces qualia donnent lieu à des états fonctionnels différents. Si au contraire on nie la possibilité d'avoir des croyances quant à nos qualia, on affirme que les états mentaux sont opaques, même pour l'individu qui a ces états, ce que les défenseurs des qualia ne sont pas prêts à affirmer.

À l'inverse, si on préfère dire que le golem, pour qu'il soit isomorphe fonctionnellement à l'agent « normal », a lui aussi des croyances qualitatives qui sont dans son cas non-fondées, alors nous devons aussi être sceptiques face à nos propres états mentaux (peut-être sommes-nous aussi des golems?). Ainsi le cas de l'absence des qualia est soit un faux problème (les croyances qualitatives que nous avons font que nous ne sommes pas isomorphes aux golems), soit il est un problème sceptique qui détruit toute la pertinence d'une enquête sur nos états mentaux.

Shoemaker nie la possibilité des qualia absents, mais ce refus n'est pas rare chez les fonctionnalistes. Toutefois, rares sont ceux qui admettent la possibilité de

l'inversion des qualia. C'est pourtant ce que fera Shoemaker. Il admet la possibilité logique de l'inversion du spectre (qui lui fera remettre en question la portée explicative du fonctionnalisme), mais il ne croit pas qu'il y ait de telles inversions dans les faits.

3.1.3 Inversion intrapersonnelle ou inversion interpersonnelle

Dans la première partie de « *The Inverted Spectrum* » (1982), il apporte une précision pertinente pour le débat de l'inversion du spectre, en distinguant l'inversion intrasubjective de l'inversion intersubjective. La description habituelle du problème de l'inversion du spectre visuel repose sur l'idée qu'il est impossible pour un agent X de savoir ce qu'un agent Y sent vraiment lorsqu'il décrit son expérience perceptuelle. Ce gouffre entre les deux agents pourrait être qualifié d'intersubjectif. Shoemaker amorce son article en examinant la proposition faite par Wittgenstein.

Consider this case: someone says « I can't understand it, I see everything red blue today and vice versa. » We answer « it must look queer! » He says it does and, e.g. goes on to say how cold the glowing coal looks and how warm the clear (blue) sky. I think we should under these or similar circumstances be inclined to say that he means by the words « blue » and « red » what we do as he always used them as we do.⁴⁵

On retrouve ici le problème de l'inversion du spectre, non pas entre différents individus, mais à l'intérieur de l'expérience d'un seul et même individu. Shoemaker utilise cette proposition car il croit qu'en distinguant les deux types d'inversion on se retrouve nécessairement à embrasser les deux. Ainsi, son argument visera dans un premier temps à montrer que l'acceptation de l'inversion intrasubjective mène nécessairement à l'acceptation de l'inversion intersubjective. Il est important de noter qu'il n'attribue pas cet argument à Wittgenstein et qu'il ne croit pas que ce dernier ait envisagé ce glissement entre les deux types d'inversion. En effet, Wittgenstein n'aurait fort probablement pas accepté l'inversion intersubjective car elle présuppose une espèce de langage privé, langage réfuté dans les « *Investigations philosophiques* ».

Le point de départ de Shoemaker est l'affirmation wittgensteinienne de la possibilité logique de l'inversion intrasubjective. Si on accepte ce genre

⁴⁵ voir Rhees (1968) p.284 cité dans Shoemaker p.358.

d'admission, Shoemaker croit que l'on doit accepter l'inversion intersubjective. Une option serait de rejeter l'inversion intrasubjective: c'est l'objection vérificationniste, qui consiste à dire que rien d'anormal n'aurait en fait vraiment lieu puisque nous ne sommes pas en mesure de prouver que nous sommes dans une situation anormale (l'inversion étant indétectable). Cette objection n'est plus vraiment prise au sérieux (bien que Shoemaker se servira dans la partie épistémologique de son article d'une version Frege-Schlick du vérificationnisme), mais même si elle l'était, Shoemaker croit que l'on observerait des différences (ces différences seraient probablement visibles dans des jugements de similarité entre deux couleurs -le gazon « rouge » pour l'agent inversé serait associé par ce dernier à une couleur proche de celle d'une orange ce qui révélerait une différence profonde entre ces jugements de couleur et ceux des agents « normaux »⁴⁶). Ainsi, il est facile d'imaginer que ce que Wittgenstein raconte nous arrive ou arrive à quelqu'un d'autre, et on peut imaginer des comportements non verbaux qui illustreraient cet état (comme la chaleur associée au rouge ou le froid au bleu)⁴⁷.

Si nous acceptons les cas d'inversion intrasubjective -c'est à dire que *pour un seul et même agent* ce qui apparaissait rouge à un moment t_1 apparaît vert à un moment t_2 - nous devons aussi accepter qu'il y a inversion intersubjective, car si l'expérience de cet agent était la même que pour les autres agents à t_1 , elle serait radicalement différente de celle des autres à t_2 . La perception de l'agent change par rapport à sa propre expérience passée, mais elle change aussi par rapport à

⁴⁶ La définition de ce qu'est un agent « normal » est une question fort problématique dans les sciences de la couleur. Cette définition ne sera pas étudiée ici, mais elle pourrait fort bien influencer les arguments anti-fonctionnalistes. Ainsi, s'il n'y a pas une « normalité perceptuelle », ne sommes-nous tous pas des agents inversés ou au moins déviants par rapport à tous les autres individus? Si c'est le cas, ou bien le fonctionnalisme semble voué à l'échec si les structures fonctionnelles diffèrent d'un agent à un autre (de la même manière que les structures neurales diffèrent d'un agent à un autre) ou bien il est condamné à particulariser ces explications pour des sous-communautés d'agents percevants partageant des structures communes; Ce sera l'avenue privilégiée par Lewis qui se sortira de l'impasse de l'inversion du spectre, en disant que l'agent inversé et l'agent « normal » ont des qualia différents ET des structures fonctionnelles différentes, chacun appartenant à une communauté différente fonctionnellement unique. Pour plus de détails quant à la notion d'observateur normal, voir Hardin (1988).

⁴⁷ Hardin (1988) utilise un argument analogue qui sera critiqué par Levine, qui affirme que l'on pourrait aussi imaginer une inversion des qualia associés aux qualia chromatiques (ce débat a été examiné dans la section 2 et sera réexaminé sous un autre angle dans la section 5 de ce mémoire).

l'expérience des autres agents. S'il y a inversion intersubjective lorsqu'il y a inversion intrasubjective, pourquoi n'y aurait-il pas d'inversion intersubjective *simpliciter*? Ainsi, à ce stade-ci on ne semble pas être en mesure de rejeter la possibilité logique qu'il y ait des agents dont l'expérience sensible est et a toujours été inversée malgré notre ignorance de cet état. L'argument de Shoemaker est étrange: l'utilisation habituelle de l'inversion du spectre vise soit sa possibilité logique, soit sa détectabilité empirique. Si on se préoccupe de sa possibilité logique, la question est de savoir s'il y a un problème ou non, alors que si on se préoccupe de sa détectabilité empirique la question est plutôt de savoir si le fonctionnalisme répond à ses objectifs explicatifs ou non. L'argument métaphysique de Shoemaker semble glisser d'un pôle à l'autre. On dirait qu'il veut que l'inversion soit possible (argument *a priori*), mais il ne garantit sa possibilité qu'en prouvant sa détectabilité (argument *a posteriori*) par l'agent lui-même, puis par les autres. Qui plus est, la possibilité logique de l'inversion intrasubjective (garantie étrangement par un fait empirique) fonde la possibilité logique d'une inversion intersubjective (non garantie par des faits empiriques) sans inversion intrasubjective préalable. Le plus étrange, c'est que c'est la possibilité d'une inversion *détectable* qui garantit une inversion (intersubjective) *indétectable*. Cet enchaînement entre les types d'inversion est décrit très clairement dans « *Functionalism and Qualia* » (1980, p.259):

If we did not think that we could have these kinds of evidence of intrasubjective spectrum inversion, I think we would have no reason at all for thinking that spectrum inversion of any sort, intrasubjective or intersubjective, is even logically possible. To claim that spectrum inversion is possible but that it is undetectable even in the intrasubjective case would be to sever the connection we suppose to hold between qualitative states and introspective awareness of them (between them and the qualitative beliefs to which they give rise), and also their connections to perceptual beliefs about the world and, via those beliefs, to behavior.

Ainsi, Shoemaker croit que c'est le lien nécessaire entre les états qualitatifs et la conscience introspective de ces états, et de la modification du comportement que cette conscience occasionne, qui garantit la détectabilité de l'inversion du spectre intrasubjective. Toutefois, tout miser sur un changement comportemental causé par une croyance à propos de nos propres qualia semble être un pari dangereux, car on peut facilement imaginer un cas où une croyance relative à nos qualia ne provoquera

pas de changement comportemental. La question devient alors de savoir si toute croyance offre nécessairement une condition de changement comportemental. Si on peut imaginer que deux croyances face à des énoncés propositionnels ne se traduiront pas par des différences comportementales, il est concevable *a fortiori* que ce sera aussi le cas face à des croyances sur les qualia⁴⁸.

Si on lui accorde le bénéfice du doute à propos de cette apparente incohérence, on peut dire provisoirement que Shoemaker nous oblige à soutenir autant l'inversion intrasubjective qu'intersubjective. Toutefois, il examine différentes possibilités de réponse.

Shoemaker examine différents arguments qui ont été avancés contre la possibilité logique de l'inversion intersubjective. Il montre aussi comment on peut dépasser ces critiques.

3.1.4 1^{ère} objection à la possibilité d'inversion du spectre: différence qualitative.

La première objection vise le statut de la ressemblance et de la différence entre différents états qualitatifs. Une manière de rejeter l'inversion intersubjective est de dire que, en fait, les états qualitatifs ne sont comparables qu'intrasubjectivement. Ainsi, il n'est pas pertinent de dire que l'expérience d'un agent, dans ce cas-ci, l'aperception chromatique, diffère de celles des autres agents, car de toutes manières, toute son expérience diffère de celle des autres agents, ses qualia étant totalement irréductibles à ceux des autres. Le problème de cette avenue est que son utilisation élimine toute possibilité de connaître ou de comprendre la vie mentale des autres agents, leurs états qualitatifs étant radicalement différents des nôtres. En fait, il serait même injustifié d'essayer d'évaluer le degré de ressemblance ou de différence entre les états qualitatifs de deux agents⁴⁹.

Shoemaker concède que la description qu'il offre du problème n'est pas une avenue que Wittgenstein aurait vraisemblablement empruntée. Bien que

⁴⁸ Je dis *a fortiori* car j'assume que certains qualia sont difficilement exprimables en mots et qu'il n'est donc pas clair qu'on puisse y puiser des croyances exprimées de manière assez raffinée pour être différenciables de manière comportementale.

⁴⁹ Cette inscrutabilité est examinée aussi dans « *Phenomenal Character* » (Shoemaker, 1994) où le problème est évité en utilisant une espèce de disjonction relationnelle qui s'appliquerait à tous. Ainsi,

Wittgenstein aurait peut-être accepté que la possibilité de l'inversion intrasubjective entraîne d'éventuelles inversions intersubjectives, il n'aurait probablement pas accepté qu'il puisse y avoir une inversion intersubjective sans cas d'inversion intrasubjective. Il admet lui-même le problème que cela crée. Shoemaker cite Wittgenstein dans Rhees (1968):

We said that there were cases in which we should say that the person sees green what I see red. Now the question suggests itself: if this can be so at all, why should it not always be the case? It seems, if once we admitted that it can happen under peculiar circumstances, that it may always happen... This is a very serious situation. (p.360)

Shoemaker souligne que Wittgenstein n'aurait probablement accepté que certains types d'inversion intrasubjective, et qu'il n'est pas clair que ces types d'inversions donneraient vraiment lieu à des inversions intersubjectives sans inversion intrasubjective préalable. Toutefois, même si on admet cette limite aux types d'inversions acceptés par Wittgenstein, Shoemaker croit que le problème reste entier, car même si Wittgenstein n'acceptait que les cas d'inversions intrasubjectives où l'agent affirme que ses perceptions lui apparaissent être étranges (ce qui correspond à une inversion décelable de manière comportementale), Shoemaker avance l'idée que, par adaptation progressive, le même agent pourrait, après de nombreuses années, s'être adapté, avoir adapté son langage, que cela ne lui semblerait plus étrange de voir le gazon « rouge », et que le comportement serait alors identique au nôtre même si ses qualia étaient différents. Ainsi, même le cas limité qu'accepterait Wittgenstein pourrait donner lieu à une inversion du spectre indécélable (ce qui constituerait un contre-exemple à sa critique du langage privé). Si on accepte ce genre d'inversion intrasubjective détectable au départ, mais qui n'est plus détectable par la suite, on doit accepter l'inversion intersubjective tout court.

il n'y aurait pas vraiment de solipsisme car tous posséderaient la même banque d'expériences possibles.

3.1.5 2^{ème} objection à la possibilité d'inversion du spectre: incompréhension plutôt qu'inversion.

Une autre objection à l'inversion du spectre est que même si on croyait avoir un cas d'inversion du spectre, nous ne pourrions jamais être certains que nous avons vraiment affaire à une inversion du spectre plutôt qu'à une transformation, chez l'agent supposément inversé, de la compréhension des couleurs ou de sa mémoire des couleurs que les objets « revêtaient » auparavant. Ainsi cette objection nous demande d'adopter une position sceptique face à l'existence possible d'une inversion intersubjective. Shoemaker répond à cette objection en partant de l'évidence que ce ne pourrait pas être seulement deux couleurs qui seraient inversées, car une telle inversion serait immédiatement détectée⁵⁰. Il faudrait que ce soit tout l'espace qualitatif chromatique qui soit inversé. Shoemaker croit que c'est la condition nécessaire pour toute inversion réelle du spectre visuel (voir aussi « *Phenomenal Character* », 1994, p.35-36). Ainsi le rouge serait encore proche de l'orange et le vert serait encore proche du bleu (alors que si on n'inversait que deux couleurs, on retrouverait le vert près de l'orange et le rouge près du bleu, ce qui détonnerait de manière évidente). Shoemaker propose qu'en prenant l'agent que l'on suppose inversé (mais qui pourrait peut-être souffrir de problème langagier face à son utilisation des termes de couleurs) et en imaginant que l'on peut faire varier son paysage chromatique progressivement (induire une inversion de manière neurologique) jusqu'à ce qu'il affirme voir la même chose que nous, il affirmera néanmoins que son expérience diffère radicalement de celle qu'il avait au départ. Ainsi, même si on pouvait penser qu'il ne maîtrise pas le langage de la couleur, le fait que son comportement ressemble finalement au nôtre, mais qu'il affirme que son expérience diffère radicalement de son expérience antérieure, démontre que son expérience initiale n'était pas identique à la nôtre. Ainsi des problèmes de mémoire ou de langage ne peuvent pas être invoqués en faveur d'un scepticisme face à la possibilité de cas réels d'inversion du spectre.

Plutôt que de dire qu'il faille douter de la bonne utilisation des termes, le sceptique devrait miser sur le solipsisme de l'expérience qualitative. Ainsi, nous ne

pouvons jamais être certains que ce que Jean perçoit comme étant qualitativement rouge pour lui corresponde à ma sensation du rouge. C'est le problème de l'inscrutabilité des autres esprits (*knowing other minds*). Shoemaker ne croit pas que cela soit un argument sérieux contre la possibilité de l'inversion du spectre, car son point de départ pour prouver la possibilité de l'inversion du spectre n'est pas la possibilité de l'inversion intersubjective, mais plutôt la possibilité de l'inversion intrasubjective. Ainsi, Shoemaker avance qu'il n'est pas vraiment important que le rouge « apparaisse identique » pour l'agent et pour moi (que ce soit « le même » rouge que les deux agents voient), mais bien, est-ce que le rouge apparaît encore « rouge » pour l'agent subitement inversé. Ainsi Shoemaker admet provisoirement qu'il n'est pas possible d'établir une similarité ou une différence entre les qualia de deux individus distincts, mais on doit admettre que l'on puisse établir de tels jugements quant à ses propres qualia. Ainsi, je saurais si l'inversion m'affecte et cette possibilité fait que SI par hasard mes qualia étaient analogues à ceux des autres face aux mêmes stimuli, ils sont maintenant différents.

Shoemaker souligne que les couleurs n'ont pas besoin d'apparaître phénoménalement identiques pour deux agents. L'important est qu'il y ait similarité intentionnelle entre les deux agents. Ainsi je n'ai pas besoin de savoir exactement *what it feels like* pour Jean de voir rouge, mais bien de savoir qu'est-ce que Jean veut dire lorsqu'il utilise le mot « rouge ». L'inversion du spectre n'a pas besoin d'une similarité phénoménale entre les agents « normaux », mais d'une similarité intentionnelle dans leur usage des mots de couleur. Il ne faudrait pas voir là, selon Shoemaker, un scepticisme face à notre capacité de connaître les autres esprits, mais bien un certain scepticisme face à la capacité de passer d'une similarité intentionnelle à une similarité phénoménale. Je suis en mesure, selon Shoemaker, de savoir si l'autre agent veut dire la même chose que moi, mais je ne suis pas en mesure de savoir si ce qu'il dit correspond exactement à mon expérience phénoménale de la chose. Seule cette similarité intentionnelle est nécessaire pour

⁵⁰ Ces deux couleurs inversées apparaîtraient incongrues dans une palette continue des couleurs.

l'inversion du spectre, l'inversion intersubjective étant de toute manière fondée sur la possibilité de l'inversion intrasubjective.

3.1.6 3^{ième} objection: possibilité concrète d'une inversion

La dernière objection à la possibilité de l'inversion du spectre est une objection empirique. Comme le disait Shoemaker plus tôt, pour qu'une inversion du spectre soit intéressante, tout l'espace qualitatif doit être translaté sinon une teinte ressortirait immédiatement parmi les teintes avoisinantes. L'objection empirique consiste à dire que l'espace chromatique n'est pas uniforme, et qu'il n'est donc pas possible de l'inverser sans que certaines couleurs apparaissent maintenant étranges dans leur « voisinage » chromatique. Shoemaker répond, avec raison d'après moi, que de tels arguments ne sont pas vraiment pertinents pour nier la possibilité « métaphysique » d'inversion du spectre, car on pourrait imaginer une créature qui aurait un espace chromatique « inversable ». À moins que le fonctionnalisme ne veuille se limiter aux humains, il doit trouver un moyen conceptuel de découvrir les cas d'inversion dans d'autres « systèmes » (ou de prouver son impossibilité), même si le « système » humain ne semble pas, selon certains, être « inversable »⁵¹.

Ayant répondu à quelques grandes critiques faites contre la possibilité métaphysique de l'inversion du spectre, Shoemaker se donne alors les coudées franches pour élaborer une éventuelle fonctionnalisation des qualia. En fait, Shoemaker croit que, même si les critiques face à l'inversion du spectre étaient concluantes, la dissolution du problème n'éclairerait pas plus d'autres problèmes. Ainsi il est probablement plus utile d'accepter provisoirement l'inversion du spectre pour voir quelles conclusions nous sommes en mesure d'en tirer, plutôt que de rejeter le problème d'emblée et de ne rien apprendre du tout. Cette démarche est d'autant plus intéressante que, comme nous le verrons bientôt, Shoemaker ne croit pas en l'existence réelle de l'inversion du spectre. Toutefois, son admission de la possibilité logique lui fournira ce qu'il croit être une définition plus robuste (parce

⁵¹ De toutes manières, même les résultats empiriques divergent sur la question. Ainsi, même si Hardin (1988) croit avoir de bonnes raisons pour croire que l'inversion serait nécessairement détectable chez l'homme, Rümelin (1996) croit qu'une altération dans les photo-pigments constitueraient un cas d'inversion possible (et probablement réel chez certains individus) qui serait indétectable.

que plus flexible) du fonctionnalisme, que l'inversion du spectre se montre éventuellement réelle ou non⁵².

3.2 Le nouveau fonctionnalisme de Shoemaker

3.2.1 Parallèles vérificationnistes

Dans le reste de son article, Shoemaker ne remet plus en question la possibilité de l'inversion du spectre (sauf brièvement à la fin de son article), mais s'attaque plutôt à voir comment le fonctionnalisme (du moins, la variante qu'il en offre) peut y survivre.

Pour que le cas de l'inversion du spectre soit intéressant, il doit pouvoir se produire de manière intersubjective. Postuler qu'il n'est qu'intrasubjectif nous place sur une pente glissante qui nous mène jusqu'au solipsisme. Une manière de garantir la possibilité d'inversion intersubjective est de décrire comment il serait possible de juger de la similarité ou de la différence entre des états instanciés chez des individus distincts. Toutefois, à la manière d'un vérificationniste comme Schlick, Shoemaker ne croit pas que l'on puisse comparer les états qualitatifs entre des individus. Ainsi, les jugements de similarité qualitative n'ont de sens qu'intrasubjectivement car seulement notre comportement peut nous éclairer à propos de nos états qualitatifs, alors que le comportement des autres ne nous apprend rien de concluant sur leurs propres états. Mais est-ce que cela va jusqu'à dire que nos jugements visant ce qui se passe dans la tête des autres agents n'ont aucune valeur? Non, car nous pouvons nous prononcer sur l'intentionnalité présentée par le comportement de l'agent. Ainsi, même si je ne sais pas ce que l'autre perçoit vraiment lorsqu'il affirme voir rouge, je sais qu'il veut dire qu'il voit rouge, même si l'aperception attachée à son affirmation diffère de la mienne.

Le vérificationnisme ne niant pas la validité des jugements de similarité intentionnelle, les discours de perception restent pertinents, mais non pour juger de la similarité qualitative entre deux agents. Un des buts de Shoemaker est de protéger

⁵² Voici comment Shoemaker décrit son pari (à la Pascal): « *I conclude that there is no fundamental epistemological or metaphysical problem here which is solvable on the assumption that our experience is not invertible but not solvable on the assumption that it is. So it will do no harm-and will make exposition easier-to assume from now on that it is.* » (Shoemaker, 1982, p.368)

une position comme celle de Frege et Schlick car il s'en servira pour fonder son fonctionnalisme. Ainsi, il croit que, même si la similarité qualitative n'est valable qu'intrasubjectivement, le fait que les états qualitatifs engendrent des croyances à propos de ces mêmes états rend possible la comparaison de croyances portant sur des états qualitatifs. Un individu ayant une expérience qualitative « stable » (i.e. quelqu'un dont l'expérience qualitative ne fluctue pas de manière étrange -un exemple d'expérience instable serait le témoignage de certains utilisateurs du viagra qui rapportent percevoir un décalage vers le bleu dans leur expérience chromatique) en vient à *croire* que son expérience qualitative des couleurs correspond à des propriétés objectives des objets. Ces croyances sont comparables intersubjectivement car elles se révèlent de manière comportementale (linguistiquement ou non). Ainsi comme il semble y avoir un lien stable entre notre expérience et nos croyances portant sur ces expériences, et comme nos croyances sont comparables intersubjectivement, le quale semble être, dans une certaine mesure fonctionnalisable.

3.2.2 Réalisation physique

Shoemaker trace ensuite la manière de pouvoir juger de la similarité de croyance. En effet, la croyance que nos jugements de couleur correspondent à des propriétés objectives des objets, n'est pas suffisante pour garantir cette objectivité. Cette objectivité doit être fondée sur quelque chose d'autre que cette même croyance. La première étape adoptée par Shoemaker sera de définir les qualia comme les propriétés des états sensoriels qui ne sont comparables que qualitativement. En tant que propriétés, les qualia sont donc des universaux. L'analyse fonctionnelle de Shoemaker commence donc en disant que le quale Q1 et Q2 sont similaires s'ils génèrent, dans l'individu qui vit ces expériences, des croyances ou des comportements relativement similaires. Nous ne sommes ici qu'au niveau intrasubjectif. Toutefois rien ne nous empêche d'affirmer que Q1 et Q2 appartiennent à deux individus distincts. On pourrait alors dire que l'agent ayant Q1 a une expérience similaire à l'agent ayant Q2 si les croyances associées sont similaires. Ceci n'est pas encore satisfaisant. Shoemaker ajoute que pour que l'on

puisse juger de la similarité de l'expérience de deux agents, on doit se servir des réalisations physiques de ces qualia. Ainsi, voici comment Shoemaker définit l'inversion fonctionnellement. Il définit les qualia comme une disjonction des propriétés des réalisations physiques. En assumant que je suis un agent « normal » on dira que Q-VR sera le quale associé à ma perception de vert et que Q-RV sera le quale associé à ma perception de rouge⁵³. En partant de la prémisse que l'inversion du spectre est possible, on dira que, devant une tomate, j'ai le quale Q-RV alors que l'agent inversé a le quale Q-VR. Si nous réussissons à définir fonctionnellement les qualia en termes de leurs rapports de similarité intentionnelle, nous pourrions alors utiliser les qualia dans une définition fonctionnelle du genre: « une paire de qualia dont un membre caractérise les perceptions de vert alors que l'autre caractérise les perceptions de rouge ». Même si les qualia sont difficilement comparables, on pourra néanmoins attribuer ces paires de qualia aux membres d'une même espèce. Il s'agira ensuite de savoir comment cette paire est réalisée physiquement dans une espèce donnée. Plusieurs réalisations physiques de Q-VR sont possibles, et la disjonction de ses réalisations physiques correspondra à la définition fonctionnelle de Q-VR. Ceci implique que le fonctionnalisme attribuant à des Martiens les mêmes états mentaux que nous doit être rejeté: seules deux créatures relativement similaires physiquement peuvent avoir des expériences comparables⁵⁴.

Shoemaker (p.376) offre une définition formelle des disjonctions acceptables. Nous ne nous attarderons pas sur cette définition car elle n'affecte pas réellement le sort du fonctionnalisme de Shoemaker. Le but est d'assigner aux qualia une réalisation physique. Ainsi, l'inversion resterait possible mais elle serait sujette à des descriptions fonctionnelles car, en théorie, on pourrait toujours découvrir quelle

⁵³ Shoemaker utilise plutôt les couleurs jaune et bleu telles qu'utilisées par Locke dans la première formulation du problème de l'inversion du spectre. Toutefois, son exemple n'exige pas l'utilisation d'une couleur plutôt qu'une autre, ce qui n'est pas le cas d'autres arguments concernant l'inversion du spectre. Comme certains arguments exigent l'emploi de certaines couleurs précises (Hardin utilise le rouge et le vert à la lumière de l'*opponent process theory*) je préfère utiliser cette même manière de décrire le problème pour standardiser le propos.

⁵⁴ « *What I am supposing about my Martians is that the sets of possible physical realizations in them of their qualia do not overlap at all with the sets of possible realizations in us of our qualia-and from this it follows that none of our qualitative states are qualitatively similar, to any degree at all, to any of theirs.* » (Shoemaker, 1982, p.375)

paire de qualia est impliquée dans tout jugement de couleur grâce à sa réalisation physique. Ainsi. Même si le vert peut-être réalisé par X1 ou X2 ou X3, si je sais que X1, X2, OU X3 sont les réalisations de Q1, je pourrai assumer que si quelqu'un dit voir vert, mais qu'il n'est pas en X1, X2, ou X3, il n'est pas dans l'état Q1. Ces réalisations physiques pourraient être fonctionnalisables en découvrant quels sont les inputs spécifiques permettant de les « activer » et quels sont les output qu'ils génèrent. Toutefois, Shoemaker se sert des réalisations physiques car les comportements ne sont pas suffisants (s'ils étaient suffisants, le problème de l'inversion ne se poserait même pas). Nous pouvons donc nous demander si la fonctionnalisation est vraiment possible s'il ne peut pas distinguer les réalisations physique de manière comportementale. Il n'est donc pas clair, du moins jusqu'ici, que la proposition de Shoemaker puisse vraiment être accommodée dans le fonctionnalisme. Décrire cette thèse comme étant physicaliste serait peut-être plus juste. De manière provisoire on peut dire que le physicalisme est accessoire : une manière de comprendre Shoemaker serait de dire que le fonctionnalisme est la théorie de l'esprit la plus efficace, mais que l'inversion n'est détectable qu'en empruntant une voie plus physicaliste. Accordons le bénéfice du doute à Shoemaker : de manière minimale il montre que l'inversion du spectre est détectable. Il reste à prouver que cette détection puisse se faire fonctionnellement ce que Shoemaker ne fait pas vraiment dans son article.

3.2.3 Conclusion

La conclusion de Shoemaker est épistémologique. il croit que, empiriquement, en découvrant ces disjonctions, nous pourrions affirmer qu'une inversion indétectable (qui sonnerait le glas du fonctionnalisme) n'est pas possible dans les faits. Même si cette conclusion n'est pas aussi forte que désirée, elle a au moins le mérite de se fonder sur un argument métaphysique. Ainsi, même s'il n'est pas évident que nous découvrirons ou comprendrons pleinement les mécanismes en jeu, le fait de lier qualia et réalisation physique sous forme de disjonction nous offre des espoirs de réduction partielle. Shoemaker défend son interprétation en se servant de l'argument de parcimonie. Ainsi, si nous étions incapables de découvrir les cas

d'inversion, il faudrait postuler des mécanismes pour expliquer cette capacité de dissimuler cette différence. On réussit difficilement à imaginer l'utilité de tels mécanismes complexes. Or, d'un point de vue évolutif, de tels mécanismes doivent être explicables.

Par contre, le supposé fonctionnalisme de Shoemaker reste litigieux. Le fonctionnalisme en tant qu'il décrit les structures en termes d'enchaînements causaux, de fonction, d'input et d'output, a une certaine visée d'élimination. Même si ce ne sont pas tous les genres de fonctionnalisme qui demandent une réalisation multiple au travers de différentes structures analogues, les modèles doivent tout de même être réductibles à une espèce de carte ou de programme. Or le fait que Shoemaker concède l'aspect mystérieux et ineffable des qualia empêche la quasi-réduction fonctionnaliste. En d'autres mots, si on espère pouvoir traduire toute l'activité mentale sous forme d'un programme; comment programmer ou introduire les « mystérieux » qualia dans cette description? C'est l'argument anti-fonctionnaliste contre ce genre de description. En quelque sorte, Shoemaker croit que ce n'est pas grave car ces propriétés mystérieuses seraient attachées à des réalisations qui, elles, sont fonctionnalisables.

Peut-être trouvons-nous ici une manière de rescaper le fonctionnalisme. De la même manière que les phénomènes indéterminés au niveau quantique peuvent être quantifiés au niveau de la physique classique (bien sûr ceci n'est qu'une analogie grossière), l'esprit pourrait aussi être compris à différents niveaux explicatifs. Ce que Shoemaker nous offre serait alors un niveau différent de description: la question reste de savoir si ce saut entre différents niveaux explicatifs peut être compatible avec le fonctionnalisme.

Outre cette question de nomenclature, il y a aussi un problème lié au fondement de l'argument de Shoemaker. Il est important de noter que la position de Shoemaker quant au statut de la couleur offre peut-être la meilleure manière d'attaquer ses arguments. Il défend une position réaliste de la couleur tout en reconnaissant le rôle joué par notre système cognitif dans la construction des perceptions chromatiques.

I will assume here that colors are where the contents of our visual experiences and our ordinary ways of placing them-on the surface of physical objects, or in expanses of sky or water. Grass is green, the sky is blue, ripe tomatoes are red. But reflection on the disparity between the manifest and the scientific image⁵⁵ makes inescapable that the conclusion that, to put it vaguely at first, the phenomenal character we are confronted with in color experience is due not simply to what there is in our environment but also, in part, to our nature, namely the nature of our sensory apparatus and constitution.⁵⁶

Bien que cet extrait n'appartienne pas à *Inverted Spectrum*, il constitue une illustration intéressante de la position de Shoemaker⁵⁷. Ainsi, il n'est pas clair dans *Phenomenal Character*, qu'une compréhension de la couleur comme étant réaliste est uniquement une image manifeste de la couleur. Shoemaker croit que la position réaliste correspond aussi à l'image scientifique. Les couleurs sont dans le monde, mais nos qualia de couleurs sont dans nos têtes. Pour appuyer la seconde idée, il en vient à citer Hardin (1988) qui avance l'idée que la distinction entre couleurs primaires et secondaires appartient au système visuel et non à une distinction entre les propriétés physiques des objets. Le problème est que Hardin va plus loin que ne le révèle Shoemaker: il est anti-réaliste à propos des couleurs. Les couleurs ainsi que les qualia sont dans la tête. Il semble donc y avoir une légère incohérence entre la thèse défendue par Shoemaker et l'argument d'autorité dont il se sert pour la défendre. Comme cette tension ne se retrouve pas dans *Inverted Spectrum*, nous ne nous y attarderons pas plus. Il suffit ici de dire que la conception de la couleur favorisée par Shoemaker devra soit se défendre sans utiliser des arguments provenant d'une thèse anti-réaliste, soit renier le réalisme de la couleur. Ce reniement ne peut pas se faire sans heurts car la disjonction fonctionnelle recherchée par Shoemaker est plus cohérente avec une position réaliste.

⁵⁵ La distinction entre *manifest* et *scientific* correspond à l'idée décrite par Sellars 1963. L'image manifeste est l'image que nous avons d'un objet lors de notre vie de tous les jours (ex.: la pomme est un objet sphérique avec une surface uniforme et continue), alors que l'image scientifique est l'image « améliorée » que nous avons d'un objet, une image approfondie mais abstraite fondée sur notre savoir scientifique (ex.: la pomme n'a pas une surface continue).

⁵⁶ Shoemaker 1994 p.24

⁵⁷ Plusieurs passages manifestent sa position réaliste quant à la couleur: « If we insist on saying that the phenomenal character really belongs to experiences or sensations, it is hard to avoid the conclusion that our sense-experience systematically misrepresents its objects in the environment-that it represents them as having features that in fact belong to the experiences themselves (...) but this seems, on reflection, to be unintelligible. » (Shoemaker, 1994, p.25)

« I think that our color concepts are, for good reasons, more « objective » than our concepts of flavors. » idem p.32

3.3 Critique de Levine

3.3.1 Introduction

Joseph Levine est un de ceux, comme nous l'avons vu dans la section 2.1 de ce mémoire, qui croient que l'intuition pro-qualia⁵⁸ doit être prise plus au sérieux que les modèles qui n'en tiennent pas compte, comme la plupart des modèles fonctionnalistes.

Dans « *Absent and Inverted Qualia Revisited* » (1988), Levine critique les arguments de Shoemaker. Comme l'indique le titre de l'article, Levine va examiner l'argument de Shoemaker contre la possibilité d'absence de qualia (exemple des zombies ou des golems) et son argument pour la possibilité d'inversion du spectre. Levine admet que la position de Shoemaker peut sembler à première vue séduisante. Shoemaker semble vouloir préserver les qualia (ménageant ainsi l'intuition pro-qualia) tout en l'intégrant d'une manière indirecte dans une explication fonctionnaliste. Toutefois, Levine ne croit pas que Shoemaker ménage vraiment une place aux qualia, ou du moins pas aux qualia défendus par les « pro-qualia ».

Levine décrit la démarche de Shoemaker comme étant ce qu'il appelle une déviation du « fonctionnalisme pur »⁵⁹. Selon lui le fonctionnalisme pur doit rejeter la possibilité de l'inversion du spectre. Toutefois, Shoemaker croit pouvoir fournir une explication fonctionnelle qui lui laisse une place. Si le modèle de Shoemaker semble plus attrayant que le fonctionnalisme pur, c'est parce que, d'une part, comme nous l'avons montré dans la section précédente, Shoemaker croit que nous ne

⁵⁸ « *The absent and inverted qualia hypotheses are thought experiments which give concrete expression to what I will call, following the Churchlands, the « pro-qualia » intuition. This is the intuition that there is something special about conscious mental life that makes it inexplicable within the theoretical framework of functionalism, and materialism more generally. The pro-qualia intuition certainly constitutes one of the major obstacles to a fully adequate resolution of the mind/body problem.* » (Levine, 1988, p.272)

⁵⁹ Levine ne fait pas de distinction entre les différents types de fonctionnalisme. En fait il ne distingue même pas le fonctionnalisme empirique et analytique. Peut-être est-ce là la plus grande faille de ces arguments anti-fonctionnalistes. Voici comment il définit le fonctionnalisme : « *Functionalism, to repeat, is the position that type-identity conditions for mental states can be specified in terms of their relations to inputs, outputs, and other mental states.* » (Levine, 1988 p.273). Bien que cette caractérisation ne soit pas fautive, elle ne reflète pas la complexité de cette position et ces arguments se révéleront être inefficace contre certaines formes de fonctionnalisme.

pouvons pas réfuter la possibilité d'inversion intrasubjective, inversion qui mène nécessairement, selon Shoemaker, à la possibilité d'inversion intersubjective. Et d'autre part, ce nouveau fonctionnalisme, laissant une place aux qualia, semble moins contre-intuitif que le fonctionnalisme pur.

Levine cherchera à montrer que le « nouveau » fonctionnalisme de Shoemaker n'est pas vraiment différent du fonctionnalisme pur, et que nous devons nous demander pourquoi on devrait adopter le fonctionnalisme de Shoemaker qui se dit plus conciliant face au qualia, mais qui finit en fait par devoir l'éliminer comme le fait le fonctionnalisme pur. L'argument de Levine se fait en trois temps. Il montre d'abord que l'on ne peut pas diviser les problèmes d'absence de qualia et d'inversion du spectre comme espérait le faire Shoemaker. Cet examen montre que les raisons pour rejeter l'absence de qualia ne sont pas cohérentes avec la description du problème de l'inversion du spectre. Il démontre ensuite que le fonctionnalisme de Shoemaker n'accommoder pas vraiment l'intuition pro-qualia. Enfin, il illustre qu'il n'y a pas un réel enchaînement logique entre la possibilité de l'inversion intrasubjective et la possibilité de l'inversion intersubjective. Je m'attaquerai d'abord à cette dernière partie car elle me semble plus facile à examiner de manière autonome, alors que les deux premières parties sont liées et demandent une exposition plus détaillée.

3.3.2 Confusion entre différents types d'inversion chez Shoemaker

Levine croit que la thèse de Shoemaker est ou bien indiscernable du fonctionnalisme pur (ce qui signifie que la prétention pro-qualia de Shoemaker n'est pas réelle) ou bien incohérente avec sa critique de l'argument de l'absence du quale. Toutefois, dans la dernière section de son article, Levine souligne un autre problème fondamental dans l'argument de Shoemaker. Sa caractérisation fonctionnelle du quale dépend de la possibilité de l'inversion du spectre. Or, l'argument de Shoemaker défendant la possibilité d'inversion du spectre est fondé sur l'idée que, comme nous devons accepter la possibilité d'inversion intrasubjective, nous devons aussi accepter la possibilité d'inversion intersubjective. Toutefois, et je crois que la critique de Levine est ici évidente, on ne parle pas du même genre d'inversion:

comme je l'ai souligné dans la description de l'argument de Shoemaker, il semble étrange que le fait qu'une inversion (intrasubjective ici) détectable soit possible mène nécessairement à la possibilité d'une inversion (intersubjective) indécélable.

Shoemaker se trompe dans sa caractérisation des inversions car si l'inversion intrasubjective est détectable, l'agent inversé n'est pas isomorphe fonctionnellement à ce l'agent qu'il était avant l'inversion. Après tout, s'il détecte un changement, c'est qu'il détecte une différence fonctionnelle. Dans le cas d'inversion intrasubjective, il n'y a donc pas d'isomorphie fonctionnelle (et l'inversion intrasubjective ne constitue pas un contre-exemple au fonctionnalisme). Toutefois, si, comme Shoemaker l'affirme, l'inversion intersubjective est possible (et qu'elle apparaît être indécélable), il y aurait alors isomorphie fonctionnelle entre les deux agents. Or, Shoemaker n'explique pas comment on peut inférer de la possibilité d'une *différence* fonctionnelle chez *un* même agent, une *isomorphie* fonctionnelle entre *deux* agents. Shoemaker nous exhorte à accepter la possibilité d'inversion intersubjective à cause de la possibilité d'inversion intrasubjective, mais ces deux inversions ne sont pas comparables (ou, du moins, Shoemaker ne montre pas comment le saut doit se faire). Qui plus est, Shoemaker croit que l'on peut défendre le fonctionnalisme contre la critique apportée par l'inversion du spectre en fonctionnalisant la ressemblance et la différence entre deux qualia. Or Shoemaker ne décrit ici cette différence qu'intrasubjectivement. Il avance que l'individu est capable de dire si deux états qualitatifs sont relativement similaires ou non. Pour se sortir de ce solipsisme et pour rendre la comparaison d'états qualitatifs appartenant à deux individus possible, Shoemaker propose ensuite que ce soient les ressemblances ou les différences entre des états *physiques* qui déterminent la ressemblance ou la différence entre des états qualitatifs. Comme nous sommes capables de distinguer des états qualitatifs intrasubjectivement, nous serons probablement capables d'identifier ces états qualitatifs à des réalisations physiques particulières. Puis nous pourrons comparer ces réalisations particulières intersubjectivement, pour déterminer s'il y a ressemblance ou différence qualitative intersubjectivement, mais comme le dit Levine, « *there is no reason to believe that from the fact that two physical states play*

different functional role within one subject, and therefore are qualitatively different, one can infer that they are also qualitatively different when they play the same functional role in two different subjects. »(Levine, 1988, p.285). Shoemaker n'explique jamais comment on peut effectuer ce saut, et il ne réussit donc pas à montrer pourquoi l'inversion du spectre peut exister tout en ne constituant pas une menace au fonctionnalisme.

Bien sûr l'argument de Levine dépend de l'asymétrie fonctionnelle dans les cas d'inversion intrasubjective. Toutefois, Shoemaker semble pouvoir éviter cet écueil, quand il propose une adaptation progressive de l'expérience de l'agent au fil des années qui suivent son inversion momentanée⁶⁰. Ainsi, après plusieurs années il ne dira plus qu'un ciel rouge est « étrange ». On peut postuler que s'étant adapté à sa « nouvelle » expérience qualitative, il sera fonctionnellement isomorphe à ce qu'il était avant l'inversion, même si son expérience qualitative est différente de celle auparavant (avant l'inversion). Le but est de montrer que, autant dans le cas de l'inversion intrasubjective que dans le cas de l'inversion intersubjective, il peut ultimement y avoir isomorphie fonctionnelle. Par contre, cet ajout ne règle toujours pas le problème soulevé par Levine. Même si on accepte cette transformation graduelle menant à une isomorphie fonctionnelle⁶¹ intrasubjective, on ne peut pas dire qu'il y a isomorphie fonctionnelle intersubjective car il y a une différence entre deux états fonctionnellement identiques chez deux individus et deux états fonctionnellement identiques *après une période d'adaptation* chez un même individu. Le problème reste de savoir si on peut appliquer les relations physiologiques que l'on trace chez un individu sur une longue période de temps à une comparaison entre deux individus au même moment. Shoemaker n'explique pas comment on peut faire cette transition et n'a donc pas montré pourquoi on doit

⁶⁰ Voir « *Inverted Spectrum* »

⁶¹ Il n'est pas évident que l'on puisse dire que cette transformation graduelle mène vraiment à une isomorphie fonctionnelle, car le souvenir d'une expérience radicalement différente fera que l'agent inversé « adapté » saura que même s'il ne considère plus son expérience qualitative comme étant étrange, il l'a déjà jugée par le passé comme l'étant. Ainsi, on ne peut pas dire que l'état adapté est isomorphe à l'état pré-inversion car le souvenir d'une expérience différente aura sûrement des conséquences comportementales et donc fonctionnelles. Même si on postule une amnésie, on ne réussit pas à se débarrasser du problème car on doit alors se demander si on a vraiment un cas

nécessairement accepter la possibilité d'inversion intersubjective. Même si on accorde à Shoemaker le bénéfice du doute face à cette possibilité (de toutes manières, Levine croit que cette possibilité doit être prise très au sérieux), il n'est pas clair que son nouveau fonctionnalisme soit consistant avec des positions qu'il avait prises concernant d'autres débats.

3.3.3 Le scepticisme dans les cas de qualia absents et d'inversion du spectre

Dans la première section de son article, Levine explique pourquoi le fonctionnalisme de Shoemaker n'est, en réalité, pas plus conciliant face à l'intuition pro-qualia que ne l'est le fonctionnalisme pur. Shoemaker nie la possibilité de cas d'absence de qualia, car il croit que les qualia engendrent des croyances associées (je crois que j'ai un quale X). Comme les croyances qualitatives ont des impacts fonctionnels, le golem et l'agent normal ne seraient pas vraiment fonctionnellement identiques à moins de dire que le golem a aussi ces croyances, mais qu'elles ne sont pas fondées. La seule manière d'obtenir l'identité fonctionnelle est de dire que les croyances que nous avons quant à nos qualia sont peut-être non-fondées. Mais cette avenue contre-intuitive mènerait à un scepticisme face à nous-mêmes, scepticisme qui rendrait caduque toute autre discussion à propos de la possibilité de connaître et comprendre les autres esprits. Ainsi nous devons rejeter la possibilité d'absence de qualia car elle mène au scepticisme face à notre propre expérience. Pour que le golem et moi soyons isomorphes fonctionnellement, il devrait aussi avoir des croyances sur ses qualia (qui dans son cas sont fictifs), mais si nous acceptons cette idée, nous devons aussi accepter l'idée que nous sommes peut-être des golems (scepticisme qui semble inacceptable au pro-qualia). Le pro-qualia propose le cas d'absence de qualia contre le fonctionnalisme alors qu'en fait le cas d'absence de qualia nous pousse à être sceptiques face à l'existence de nos propres qualia. Le pro-qualia ne peut donc pas invoquer le cas d'absence de qualia.

Levine croit que Shoemaker ne peut pas nier la possibilité d'absence de qualia alors qu'il accepte la possibilité d'inversion du spectre, car sa position

fonctionnaliste implique aussi la possibilité d'un scepticisme dans le cas de l'inversion du spectre. S'il rejette l'absence de qualia à cause du scepticisme résultant, il devrait aussi rejeter l'inversion du spectre pour la même raison, et s'il rejette l'inversion, son fonctionnalisme n'est pas plus conciliant que le fonctionnalisme pur face à l'intuition pro-qualia.

La tâche de Levine est donc de montrer comment l'inversion du spectre telle que décrite par Shoemaker implique une forme de scepticisme. Ce scepticisme concernera encore une fois les croyances de l'agent à propos de ces états qualitatifs. Posons que R est l'état qualitatif *normalement* associé au rouge et V est l'état qualitatif *normalement* associé au vert. Si l'inversion du spectre est possible et si elle implique que l'agent normal et l'agent inversé sont fonctionnellement isomorphes, les croyances engendrées par R chez l'agent normal et par V chez l'agent inversé seront les mêmes. Si nous avons les mêmes croyances, mais que la croyance de l'agent inversé est fautive, ne doit-on pas admettre la possibilité que nos croyances soient fautes aussi? Dans ce cas, le scepticisme qui poussait Shoemaker à rejeter l'absence de qualia, devrait le pousser aussi à rejeter l'inversion du spectre car, dans les deux cas, nous semblons être poussés à un scepticisme face à nos propres états mentaux. Voici comment Levine exprime le problème:

If Sally [l'agent supposément inversé] and I both believe we are having experiences of type B and yet she is mistaken, what grounds do I have for thinking that I am not also mistaken? Thus the inverted qualia hypothesis leads to the extremely implausible conclusion that I can't know which type of qualitative character my sensation has. (Levine, 1988, p.279)

Il faut souligner que Levine n'est pas en train de nier la possibilité d'inversion du spectre, ou de dire que les arguments sceptiques sont nécessairement des réductions à l'absurde, mais bien que Shoemaker ne peut pas rejeter l'argument de l'absence de qualia parce qu'il mène au scepticisme, tout en maintenant que l'inversion du spectre est possible alors qu'elle est sujette au même type de critique. La réalisation physique du quale telle que décrite par Shoemaker ne résout pas le problème car de la manière dont Shoemaker définit cette réalisation physique (de manière disjonctive), nous ne pouvons pas savoir si je sens R et Sally V et non l'inverse. La caractérisation offerte par Shoemaker ne nous permet pas de dire qui a

vraiment quel quale. Personne ne peut vraiment savoir s'il est un agent normal ou inversé, et le scepticisme est donc maintenu. La seule manière pour Shoemaker de sauver l'inversion du spectre de ce scepticisme est de définir le quale de telle sorte qu'il soit ne pas défini disjonctivement. Mais que reste-t-il alors de la spécificité du quale? S'il y a une définition du quale en termes physique le « *what it is like* » du quale devient alors superflu. Levine souligne que cette option n'est pas nécessairement erronée en soi (bien qu'il croit personnellement que l'intuition pro-qualia soit plus forte), mais elle contredit ce que Shoemaker croyait pouvoir accomplir avec son nouveau fonctionnalisme.

Shoemaker croit pouvoir fournir un nouveau fonctionnalisme qui rende justice à l'intuition pro-qualia. Pour ce faire, il présente un fonctionnalisme qu'il croit capable de résister à l'argument de l'inversion du spectre, qu'il considère comme un genre de test-limite du fonctionnalisme. Mais Levine avance que la description du problème de l'inversion du spectre proposée par Shoemaker n'est pas cohérente avec sa critique de l'argument de l'absence de qualia. Pour rendre ses deux caractérisations cohérentes, Shoemaker devrait dire que l'inversion du spectre n'est pas possible à cause du scepticisme qu'elle implique. Toutefois, cette option effacerait la nécessité du nouveau fonctionnalisme qu'il propose. Il pourrait plutôt effacer le scepticisme en disant que le quale est défini physiquement de manière non-disjonctive. Toutefois, ceci détruit la pertinence de l'intuition pro-qualia qu'il prétend défendre. De plus une telle définition du quale efface aussi le scepticisme dans l'argument de l'absence de qualia⁶².

3.3.4 *Faussees concessions envers la position pro-qualia*

Il s'agit maintenant d'examiner plus précisément comment la caractérisation du quale que Shoemaker peut offrir pour effacer le scepticisme ne rend pas justice à l'intuition pro-qualia et relègue son fonctionnalisme au fonctionnalisme pur. Dans « *Absent Qualia Are Impossible* », Shoemaker suggère que pour éviter le scepticisme

⁶² « *So, if such grounds do exist for knowing the identity of the quale one is experiencing, even given the possibility of inverted qualia, then similar grounds must exist for knowing that one has any qualia at all, even given the possibility of absent qualia. Shoemaker can't have his cake and eat it too.* » (Levine, 1988, p.280)

dans le cas de l'absence de qualia, on pourrait adopter une position « *parochial* » (voir note 24 de ce mémoire pour une explication de ce terme). Shoemaker décrit cette alternative comme étant la position où la référence des termes mentaux serait fixée, à la Kripke, à des états physiologiques. Ainsi qualia et réalisation physiologique particulière seraient intrinsèquement liés. Le scepticisme serait donc évité car, *pour nous*, réaliser un qualia dépendrait d'une réalisation physique particulière, alors que cette réalisation physiologique ne serait pas présente chez le zombie ou le golem. La position *parochial* rend donc possible le cas d'absence de qualia, car l'être qui serait sans qualia ne serait donc pas comparable à nous, ses états physiologiques étant différents, et nous ne serions donc pas susceptibles de tomber dans un scepticisme face à nos propres états mentaux. Toutefois, dans ses remarques à propos de l'absence de qualia, Shoemaker ne semble pas prêt à adopter la position *parochial*, ou chauvine, car elle implique une position anthropocentrique face à la conscience. Un fonctionnalisme adoptant une position chauvine se doit de définir la conscience et les qualia en termes de *nos* états physiques. Ceci revient à dire que « conscience » et « qualia » ne peuvent pas être appliqués à des êtres physiologiquement différents de nous. Or Shoemaker ne veut pas réduire ces termes mentaux ainsi car, mis-à-part le chauvinisme que cela implique, une telle réduction ferait de toute entreprise cherchant à comprendre la conscience ou nos états qualitatifs une activité a posteriori et contingente, non-intéressante philosophiquement. Shoemaker semble donc vouloir rejeter la position chauvine (sans compter que ce genre de position attaque la possibilité de réalisation multiple défendue habituellement par le fonctionnalisme).

Toutefois, Levine croit que Shoemaker se doit d'endosser la position *parochial* qu'il a lui-même rejetée plus tôt. En effet, le scepticisme que Shoemaker avait identifié dans le cas d'absence de qualia, serait aussi présent dans le cas d'inversion du spectre. Voici comment Levine caractérise la position ambiguë de Shoemaker face à la position *parochial*:

Thus his position, as I reconstruct it, amounts to a rejection of parochialism when it comes to the general property of having qualitative character at all, but acceptance of parochialism when it comes to the particular property of having a qualitative character of some particular type. (Levine, 1988, p.282)

Cette position semble comporter plusieurs avantages. En rejetant le chauvinisme général, elle ne rejette pas la possibilité que d'autres créatures, aux physiologies différentes de la nôtre, puissent être conscientes. Toutefois, elle évacue la possibilité du scepticisme en disant que *nos* qualia étant réalisés par *notre* physiologie particulière ne sont pas comparables aux qualia (présents ou non) d'éventuels zombies.

Toutefois, cette position mitoyenne que Shoemaker semble devoir adopter pour éviter le scepticisme dans le cas de l'inversion du spectre, enlève toute la puissance à l'intuition pro-qualia qu'il affirme vouloir défendre. En identifiant rigidement les qualia à des états physiologiques, Shoemaker efface la pertinence de toute discussion sur les qualia. Si la seule manière d'éviter le scepticisme face à nos états mentaux est de les identifier à des états physiologiques, il est légitime de se demander pourquoi on parle des qualia dans un premier temps et non pas tout simplement d'états physiologiques. Car si on trace cette identité entre qualia et réalisation physique, il est trivial de postuler que deux êtres ont des expériences qualitatives différentes quand tout ce que l'on a à prouver c'est qu'ils sont physiologiquement distincts⁶³.

Ainsi, Levine croit que Shoemaker fait face à un dilemme. Ou bien le scepticisme qui lui fait rejeter le cas d'absence de qualia lui fait aussi rejeter le cas d'inversion du spectre, et sa position n'est pas plus conciliante à l'égard des qualia que ne l'est le fonctionnalisme pur. Ou bien il réussit à éviter ce scepticisme face à l'inversion du spectre, mais il doit alors rejeter ce scepticisme aussi dans le cas de l'absence de qualia, et ce seulement au prix d'une caricature des qualia de manière à vider toute la force de l'intuition pro-qualia que Shoemaker veut défendre. Le problème est d'autant plus important que, même si on réussissait à montrer que cette caractérisation n'implique pas vraiment une trivialisaiton des qualia (ce qui semble à première vue peu probable), le cas de l'absence de qualia étant sain et sauf parce que

⁶³ « *To avoid the skeptical argument, the parochialist has to say that all there is to the question of the identity of a particular quale is the question of the identity of its underlying physiological structure. But then the question of why anyone should care becomes compelling indeed.* » (Levine, 1988, p.283).

le scepticisme n'a plus prise sur lui, le fonctionnalisme se retrouve alors avec ce convaincant contre-exemple. Shoemaker ne peut pas ménager l'intuition pro-qualia et le fonctionnalisme en même temps. Comme il semble plus évident que Shoemaker préfère défendre le fonctionnalisme, Levine en conclut que son fonctionnalisme n'est pas si « nouveau » qu'il le prétend.

3.4 Conclusion

La position de Shoemaker est instable. On ne réussit pas à montrer comment le fonctionnalisme peut rester viable tout en ménageant une certaine spécificité aux qualia tels qu'envisagés par les pro-qualia comme Levine. Or, Shoemaker croit que sa théorie fonctionnaliste est supérieure à d'autres parce qu'elle respecte l'intuition pro-qualia. Levine montre que le fonctionnalisme de Shoemaker n'est pas différent du fonctionnalisme du sens commun et que le « nouveau » fonctionnalisme de Shoemaker ne réussit pas à se démarquer des autres thèses fonctionnalistes. Les éléments qui devraient distinguer sa position du fonctionnalisme pur, soit la concession envers la position pro-qualia, affaiblissent le caractère fonctionnaliste de sa théorie. Le fonctionnalisme semble obligé, du moins suite à l'examen de Shoemaker et à la critique de Levine, de se débarrasser des qualia ou de les associer de manière intrinsèque à une disjonction d'états physiques. Or cette deuxième option n'est pas nécessairement fonctionnalisable. Shoemaker n'a donc pas vraiment montré comment sa position pouvait être comprise de manière fonctionnelle. Il a seulement démontré comment le physicalisme pouvait révéler l'inversion. Or l'inversion du spectre se veut un argument contre le fonctionnalisme. Levine dit lui-même (voir section 4.1 de ce mémoire) que l'inversion n'est pas un argument aussi fort contre le physicalisme à cause des identités éventuelles qu'il peut tracer entre des expériences phénoménales et des états physiques.

De plus, même en accordant le bénéfice du doute à propos du fonctionnalisme de la position de Shoemaker, on doit remettre en question l'inférence qu'il essaie d'opérer entre les différents types d'inversion. Alors qu'il s'était forcé à bien distinguer l'inversion intrasubjective de l'inversion intersubjective, il trace des parallèles gratuits entre les deux types d'inversion. Une

inférence justifiée entre l'inversion intrasubjective et l'inversion intersubjective n'est pas vraiment démontrée. Cette critique de Levine est intéressante : Levine veut que l'inversion soit possible, mais ce qu'il montre ici est que Shoemaker ne réussit pas à prouver sa possibilité intersubjective. Ce volte-face apparent n'est en fait qu'une manière de montrer que Shoemaker ne comprend pas vraiment la position pro-qualia, ou ne réussit pas vraiment à l'intégrer au fonctionnalisme comme il espère le faire. Ainsi Levine croit que l'inversion reste un mystère contrairement à ce que Shoemaker affirme.

Finalement, comme le montre Levine, Shoemaker ne peut pas vraiment distinguer les cas de qualia absents de ceux de l'inversion du spectre. Sa position fonctionnaliste (si on peut vraiment la comprendre comme telle) ne peut pas rejeter le cas des qualia absents à cause du scepticisme général qu'il entraîne, tout en niant qu'un tel scepticisme est tout aussi présent dans le cas de l'inversion du spectre. Or Shoemaker cherchait à ménager une place à l'inversion du spectre, pour montrer que son fonctionnalisme accordait plus de valeur à l'intuition pro-qualia que ne le font les autres thèses fonctionnalistes. Shoemaker doit donc choisir entre accepter la possibilité des qualia absents et de l'inversion du spectre (ce qui attaque directement toute thèse fonctionnaliste), ou rejeter en bloc ces deux expériences de pensée (ce qui fait de son fonctionnalisme une thèse tout aussi anti-qualia que les autres théories fonctionnalistes). Le fonctionnalisme de Shoemaker ne semble donc ni une heureuse innovation, ni une solution plus acceptable pour les pro-qualia.

4 Cool Red

4.1 Argument de Levine

4.1.1 Introduction

La défense a priori du fonctionnalisme offerte par Shoemaker n'étant pas concluante, nous devrions peut-être en revenir aux suggestions plus empiriques. Levine croit aussi que ce type d'objection empirique mérite une attention sérieuse. Comme nous l'avons vu dans « *Absent and Inverted Qualia Revisited* » (1988) et surtout dans « *Materialism and Qualia: The Explanatory Gap.* » (1983), Levine utilise l'inversion du spectre non seulement pour montrer l'échec du fonctionnalisme, mais également pour montrer les limites du matérialisme en général. Cette attaque se fait surtout sur le plan épistémologique (alors qu'une attaque comme celle de Kripke se voulait métaphysique) et elle mérite un examen approfondi. Déjà dans le débat entourant « *Materialism and Qualia: The Explanatory Gap* » (1983), nous avons pu voir comment Levine et Hardin essaient de décrire l'inversion du spectre. En ce sens, l'examen que nous offrirons ici de « *Cool Red* » (Levine, 1991) qui se veut une réponse à certaines idées présentées dans le livre de C.L. Hardin, ici l'apôtre d'une démarche plus empirique, (« *Color for Philosophers; Unweaving the Rainbow* » 1988), est en quelque sorte la suite du débat amorcé en 1983. Toutefois, nous n'avons pas ici une simple répétition de l'argument de Levine; les arguments de Hardin ont été raffinés et la critique de Levine se doit donc d'être plus sophistiquée. De plus, nous verrons grâce à cet examen non seulement à quel point une question comme l'inversion du spectre modifie les limites entre le physicalisme et une forme ou une autre de fonctionnalisme, mais aussi la manière d'envisager les expériences de pensée en général.

4.1.2 L'expérience de pensée telle qu'envisagée par Hardin

L'objectif n'est pas d'examiner les rouages mis en œuvre dans les expériences de pensée⁶⁴, mais plutôt de montrer qu'il semble important de bien

⁶⁴ On peut trouver d'excellents examens de la question dans « *Tought Experiments* » de Roy A.

définir les limites de concevabilité lors de l'utilisation d'une expérience de pensée comme l'inversion du spectre. Il faudrait voir si la définition physicaliste offerte par Hardin n'est pas sujette à un argument basé sur l'inversion du spectre, d'une part par des arguments *a priori* comme ceux de Levine, mais aussi par des arguments *a posteriori* comme ceux de Nida-Rümelin. Toutefois, ce genre d'argument empirique ne sera pas étudié ici.

Dans sa réponse à Levine à propos du fossé explicatif, Hardin avait déjà présenté les cellules intermodales comparatrices, ces mêmes cellules sont décrites plus en profondeur dans son livre. Ce sont ces mêmes cellules dont Levine se servira pour réaffirmer l'existence d'un problème dans l'inversion du spectre. Comme le souligne Levine dans « *Cool Red* », la réponse de Hardin (face au supposé fossé explicatif entre l'aspect physique et l'aspect phénoménologique de l'expérience) est un mélange d'une théorie relativement bien établie et de spéculation pertinente. L'argument de Levine ne cherchera pas à rejeter *a priori* l'explication matérialiste de Hardin, mais bien de retourner l'argument contre lui-même, espérant ainsi montrer que même en acceptant les tactiques matérialistes, auxquelles Levine accorde ici le bénéfice du doute, le matérialiste n'arrive pas à tout expliquer de l'activité mentale.

Levine précise que l'inversion du spectre visuel n'est pas un contre-exemple aussi frappant pour le physicalisme que pour le fonctionnalisme, car le physicalisme peut se permettre d'identifier des qualia à des états neurophysiologiques (détour utilisé par Shoemaker pour essayer de rescaper le matérialisme). Toutefois, Levine croit que même si le physicalisme réussissait à échapper aux problèmes de l'inversion, il ne réussirait tout de même pas à expliquer pourquoi nos perceptions sont accompagnées d'expérience phénoménale. Nous ne voyons pas comment un simple examen d'un réseau neuronal pourrait révéler les différences entre deux expériences phénoménales d'un même objet. C'est ce que Hardin essaiera de faire en identifiant des associations intrinsèques entre nos perceptions de certaines couleurs et des jugements selon d'autres modalités. Rappelons-le, Hardin se base sur l'*opponent process theory*, théorie stipulant que les couleurs sont codées sur trois

Sorensen (1992) ou dans « *Thought Experiments in Science and Philosophy* » (eds, Horowitz and

canaux distincts, le canal rouge-vert, le canal jaune-bleu, et le canal achromatique (qui correspond aux gris). Si le canal rouge-vert est stimulé alors que les deux autres ne sont pas activés, nous avons la couleur rouge. Si le canal rouge-vert est inhibé alors que les deux autres ne sont pas activés, nous avons la couleur vert. Hardin concède que l'inversion est possible, mais seulement entre deux canaux (revoir la section 2.2). Hardin ne défend pas l'idée que l'inversion est impossible, mais bien qu'elle est *nécessairement* détectable. Une cellule comparatrice intermodale, et c'est ici que l'argument de Hardin devient plus spéculatif, relie l'état d'activation des canaux chromatiques à d'autres modalités sensorielles. Le rouge (pôle positif de son canal) est associé à la chaleur, alors que le bleu (pôle négatif de son canal) est associé au froid. L'expérience phénoménale du rouge, ne serait pas seulement l'activation du canal rouge-vert, mais aussi l'état de la cellule comparatrice intermodale.

Posons que Jane serait l'observatrice « normale »⁶⁵ alors que Dick serait l'agent inversé (ce qui veut dire qu'il aurait une expérience phénoménale verte en observant une tomate mûre). Comme il est inversé depuis la naissance, son langage ne semble pas être en mesure de nous révéler l'inversion (il a appris à appeler « vert » l'expérience phénoménale que nous aurions appelée « rouge » et vice versa). En tenant compte de la cellule comparatrice intermodale, Hardin examine trois possibilités⁶⁶.

1- Lorsque Dick regarde une tomate, il affirme que sa couleur est « froide » (alors que Jane dit qu'elle est « chaude ») à cause de l'inversion des pôles de ses canaux chromatiques. Dans ce cas-ci, il n'y a pas d'isomorphie fonctionnelle entre Jane et Dick car leurs comportements diffèrent de manière observable. Le fonctionnalisme

Massey; 1991).

⁶⁵ Il est important de répéter que cette question de normalité pourrait elle-même être mise en doute. Ainsi, Lycan (1996) présente l'idée que si on accepte la possibilité d'une inversion du spectre chez quelques individus, il faut aussi accepter la possibilité d'une inversion du spectre massive dans la population. Si la moitié de la population est inversée (ce qui est une possibilité facilement défendable une fois que l'on a accepté la possibilité de l'inversion du spectre), il faut se demander ils sont inversés par rapport à qui. Si 50% de la population est supposément déviante, rien ne peut montrer qu'ils sont vraiment déviants par rapport à une hypothétique normalité établie.

⁶⁶ Je m'inspire ici du résumé de l'argument offert par Levine (1991)p.30, mais on peut aussi se rapporter à l'argument original donné Hardin dans « *Color for Philosophers; Unweaving the*

(et par conséquent le matérialisme) est donc sauf contre l'inversion du spectre visuel.

2A- Les canaux chromatiques de Dick et de Jane fonctionnent de la même manière. Jane et Dick disent que la tomate est « chaude ». Par contre, si on accepte que Dick est inversé, cela veut dire qu'il a quand même l'expérience phénoménale de « froid » même s'il dit que la tomate est « chaude ». Ceci veut dire que ses « qualia » sont des épiphénomènes incapables de modifier le comportement. Il y aurait ensuite un dilemme quant à l'explication de son expérience. Ou bien il aurait en voyant une tomate mûre des qualia de vert et de froid, ce qui signifie qu'il n'est pas capable d'exprimer convenablement ses croyances⁶⁷, ou bien il croit vraiment que la tomate est rouge et chaude, ce qui signifie que ses expériences sont impénétrables non pas seulement pour l'agent externe, mais pour l'agent lui-même. Aucune des deux possibilités ne peut être endossée par le pro-qualia.

2B- Les canaux de Dick sont inversés AINSI que sa cellule comparatrice. En voyant la tomate il a l'expérience phénoménale du vert et l'expérience phénoménale du chaud. Il dira donc que la tomate est chaude, même s'il la *voit* verte⁶⁸. Le comportement ne serait donc pas en mesure de révéler l'inversion, et le fonctionnalisme serait donc un échec. Toutefois, Hardin trouve cette possibilité étrange.

But exactly how could phenomenal green be experienced as positive and warm and yet be green? If there is a « residue » of green which is separable from its polarity, that residue would seem to correspond to nothing in experience or imagination, yet it is only the ostensible imaginability of the hue inversion that makes its possibility intuitively plausible. (Hardin, 1988, p.138)

Levine offre trois critiques à la description du problème offerte par Hardin :

Rainbow » p.138.

⁶⁷ Hardin semble voir le quale comme étant majoritairement intentionnel ou relationnel ce qui ne correspondrait probablement pas à la description qu'en ferait un pro-qualia.

⁶⁸ Il est bien important de comprendre comment la double inversion fonctionne. Dick a d'une part sa cellule comparatrice inversée: ceci veut dire que *s'il n'était pas victime d'une inversion du spectre*, il dirait que la tomate qu'il verrait comme était phénoménalement rouge (parce qu'il n'est pas encore inversé), serait froide (car sa cellule comparatrice qui devrait associer la polarité positive chromatique avec une autre modalité positive comme la chaleur, associe en fait la polarité positive chromatique avec des sensations « négatives » comme le froid). Le rouge de la pomme (codage positif) est donc associé avec le froid (codage négatif d'une autre modalité). Le problème devient intéressant car Dick est non seulement inversé au point de vue de sa cellule comparatrice, mais aussi au niveau du spectre visuel. Donc il voit la tomate comme étant phénoménalement verte, mais il associe le vert au chaud. Son comportement ne pourra donc pas révéler l'inversion.

a) L'argument de Hardin n'est pas à proprement parler physicaliste, mais bien fonctionnaliste (comme le dit Levine, ce commentaire n'est pas tant une critique qu'une remise en contexte). b) Une inversion indétectable reste possible même si l'expérience phénoménale de la couleur semble être associée de manière intrinsèque à d'autres états. c) MÊME si Hardin a raison sur toute la ligne, le fonctionnalisme ne réussit pas à expliquer tout de la mentalité : elle ne peut pas expliquer pourquoi nous avons une expérience phénoménale. Comme je l'ai déjà souligné, la spéculation de Hardin est tout au plus optimiste, car il espère que les supposées associations entre couleurs et autres sensations sont plus que de simples associations culturelles. Levine ne pourrait pas utiliser ce genre d'argument car son anti-matérialisme rendrait sa critique paradoxale - il ne peut pas utiliser un argument matérialiste (on prouve empiriquement que l'association des couleurs est culturelle) pour montrer que le matérialiste a tort. Toutefois un physicaliste pourrait attaquer la thèse *pseudo-fonctionnaliste* de Hardin en montrant que sa théorie n'est pas adéquate pour rendre compte de l'expérience phénoménale, car les relations qu'il présente ne sont pas intrinsèques ou ne concernent pas un problème réel (c'est le chemin que pourrait peut-être emprunter un éliminativiste comme Dennett ou les Churchland en se débarrassant des qualia).

Examinons maintenant ces trois critiques en détails. Il faut souligner que seule la deuxième critique se révèle une attaque en règle contre l'argument anti-inversion de Hardin. Par contre, la première n'est qu'une question de taxonomie (qui bien qu'importante n'attaque pas la cohérence de la position foncièrement matérialiste de Hardin), alors que la troisième n'est qu'une critique générale contre le matérialisme (question qui se poserait même si l'inversion du spectre ou les cas de qualia absents étaient impossibles). Toutefois, l'examen des trois critiques reste pertinent pour notre étude, car il obligera Hardin de mieux qualifier sa position et ainsi de mieux camper son argument contre l'inversion du spectre.

4.1.3 *Le fonctionnalisme chez Hardin*

Le premier point concerne la nature réelle de la position de Hardin. Est-il physicaliste ou fonctionnaliste? Comme nous l'avons déjà vu dans le commentaire

de « *Materialism and Qualia: the Explanatory Gap* », Hardin se dit physicaliste, mais sa position ne semble pas être incompatible *a priori* avec le fonctionnalisme (voir section 2.2 de ce mémoire). Toutefois, il s'oppose au fonctionnalisme du sens commun qui, en se basant sur le sens commun pour établir la fonction de systèmes aussi variés que des robots ou des extraterrestres, fournit des explications générales et triviales à propos de la mentalité. Levine ne fait pas de distinction entre les formes du fonctionnalisme, car (du moins c'est l'impression qui ressort de sa description très grossière du fonctionnalisme) il juge que, comme le fossé explicatif concerne toute explication matérialiste, il n'est pas vraiment important de différencier les différentes variantes du fonctionnalisme.

Levine croit que l'argument de Hardin est fonctionnaliste car, en essayant d'expliquer les différences phénoménales en termes de différences physiologiques, il se base sur un *état* de la cellule comparatrice. La cellule est là pour comparer deux choses, et c'est cette comparaison qui nous apprend quelque chose. Or ce genre d'explication relationnelle est davantage l'apanage d'une description fonctionnaliste que celui d'une réduction physicaliste, car il se base sur l' *information-process* (traitement de données) d'un système. Cet appel à un genre de *software* n'est pas réductible à la mécanique physicaliste. Si Levine juge si important de placer Hardin dans le camp des fonctionnalistes, c'est qu'il croit que la position fonctionnaliste est foncièrement incapable de répondre à l'argument de l'inversion du spectre (ou à toute autre fluctuation de qualia), alors que le physicaliste « orthodoxe » peut y répondre en rejetant *a priori* l'expérience phénoménale⁶⁹.

4.1.4 D'autres cas d'inversion du spectre et le compromis de Levine

Levine se servira ensuite de ce changement de taxonomie pour attaquer l'argument anti-inversion de Hardin. En le décrivant comme fonctionnaliste (plutôt que comme physicaliste), Levine espère montrer que Hardin n'épuise pas vraiment les cas possibles d'inversion et que certains cas sont possibles et rendent toute la force à l'argument de l'inversion du spectre. Levine commencera par attaquer l'idée

⁶⁹ Levine croit que cet éliminativisme bien que contre-intuitif et erroné, est tout au moins plus cohérent que le fonctionnalisme.

que 2B, identifié plus tôt⁷⁰, était proche de l'incohérence. Une fois qu'il aura montré que 2B est une possibilité sérieuse d'inversion du spectre indétectable, il proposera deux autres cas (2C et 2D) pour montrer que l'inversion serait possible de différentes manières.

Malheureusement, son attaque de 2B révèle une limite du débat en général. Levine croit qu'il y a plus pour une couleur donnée que son rôle fonctionnel et sa réaction associée (Hardin ne le croit pas). Il y aurait un « résidu qualitatif » qui nous permet de dire que malgré toutes les inversions possibles, nous avons encore affaire à des entités monadiques et ineffables. Mais ceci reste une intuition (de la même manière que c'est une intuition qui pousse Hardin à dire qu'il n'y a pas de tel résidu). Hardin ne croit pas qu'on puisse encore dire que nous avons affaire à une expérience phénoménale de rouge, si on croit qu'elle est froide, car cela voudrait dire que le rouge pourrait être défini sans son rôle fonctionnel et sans sa réaction associée (qui dans le cas du rouge correspond à la chaleur), mais que reste-t-il alors du rouge? Levine essaie de rendre ce « résidu » de rouge plus intuitif, mais ne réussit pas vraiment. Son argument contre 2B reste une pétition de principe:

But that does not mean there could not be another kind of qualitative character, reddish, which shared some features with our reddish qualia and some (i.e. being coolish) with our greenish qualia. So long as one does not believe that occupying the relevant functional role together with possessing the appropriate polarity value adds up to a full-fledged visual qualia, one should find the idea of a qualitative residue intuitively plausible. Once we have gone that far, it is not too much farther to conceive of an inversion of just that feature. (Levine, p.33)*

Si nous ne sommes pas d'accord avec sa description du quale visuel, il est difficile de voir comment Levine peut avoir raison. Ainsi Levine nous demande plus un acte de foi que d'accepter un argument. Voyons si 2C et 2D sont plus crédibles.

Dans 2C, Levine envisage la possibilité que Jane et Dick n'aient pas de cellule comparatrice (alors que le reste de la population en aurait). Ainsi, Jane et Dick ne comprennent pas ce que les autres agents veulent dire quand ils disent que le

⁷⁰ 2B concernait le cas étrange où Dick souffrirait d'inversion du spectre et d'inversion de sa cellule comparatrice ce qui rendrait son comportement indifférenciable du nôtre. Hardin trouvait que ce cas était très proche de l'incohérence car si on peut inverser la réaction associée (chaud ou froid) et inverser l'expérience phénoménale, il faut se demander qu'est-ce qui reste vraiment de la couleur. Si on peut tout inverser sans rien changer, il faut se demander si l'objet étudié avait vraiment quelque chose de constitutif.

rouge est vaguement chaud ou le vert est vaguement froid. Par contre Jane et Dick auraient une inversion de leurs canaux chromatiques. Hardin utilisait la cellule comparatrice pour montrer qu'une inversion indétectable était impossible à cause des réactions associées à l'expérience chromatique. Mais sans la cellule comparatrice, les comportements de Jane et Dick seraient analogues, et nous ne serions donc pas en mesure de dire si Dick est inversé ou non. Serions-nous prêts à dire que Jane (qui correspondait à l'observatrice « normale » dans la description initiale) a un quale visuel radicalement différent du nôtre parce qu'elle n'associe pas les couleurs à des réactions relatives à d'autres modalités? Intuitivement, nous sommes portés à répondre négativement, et ceci implique que le fameux résidu de Levine ne semble pas être si farfelu que cela, car nous semblons être capables d'accepter un quale visuel analogue au nôtre qui ne serait pas lié à des réactions associées. *A fortiori*, si on accepte que Dick puisse être inversé et que son inversion est indétectable, cela veut dire qu'une couleur n'est pas exclusivement définie en termes fonctionnels ou de réaction associée (car nous avons déjà accepté que Jane a un quale analogue au nôtre même si elle n'a pas la réaction associée). 2C a donc un double effet. Non seulement il présente un autre cas d'inversion auquel le fonctionnaliste doit répondre, mais il désamorce la critique d'incohérence offerte par Hardin face au cas 2B, car le résidu qualitatif n'est plus une chose si étrange qu'elle ne l'apparaissait dans 2B.

L'argument de Levine est donc, comme il le dit lui-même, la réaction anti-fonctionnaliste classique car elle cherche à montrer qu'une explication fonctionnaliste se doit d'expliquer *tous* les cas d'inversion de qualia possibles, ce qu'il juge être impossible. Mais la prétention de Hardin est plus modeste, comme le souligne Levine (p.34): « *Hardin, representing the defenders of functionalism, claims that it is sufficient for his side to demonstrate that **our** qualia manifest the structural asymmetries necessary to render inverted qualia impossible.* ». Ces deux points de vues sont irréconciliables, c'est pourquoi Levine propose un compromis entre sa critique anti-fonctionnaliste classique (dont l'extension semble impossible à satisfaire) et la visée modeste de Hardin (qui semble faire fi de la réalisation multiple

qui est pourtant un dogme du fonctionnalisme). Levine croit que l'exigence pour le fonctionnaliste ne sera pas d'expliquer *tous* les cas possibles d'inversion, mais bien les cas qui sont assez proches de notre expérience commune actuelle. Son cas 2C répondrait à cette contrainte en tant qu'il ne propose pas une expérience farfelue ou totalement étrangère à l'expérience « normale », mais seulement un cas de déficience relationnelle (l'agent ne fait pas toutes les relations que l'agent « normal » effectue habituellement). Comme le fonctionnalisme de Hardin ne semble pas être en mesure de répondre à cette contrainte limitée, il échoue.

Bien que le compromis offert par Levine soit intéressant parce qu'il rend la critique pro-qualia plus généreuse, il pose tout de même problème. Je dis que ce compromis est plus généreux car la contrainte de devoir expliquer le phénomène à travers tous les mondes possibles rendait la position pro-qualia trop exigeante (il est loin d'être clair que quelque théorie que ce soit, est capable de répondre à un tel critère). Or un compromis comme celui offert par Levine limite le domaine qu'une théorie de l'esprit doit expliquer. Une théorie ne devra pas expliquer *toutes* les expériences, mais les expériences assez analogues à celles que nous connaissons en ce moment. Ce genre de compromis occasionne plusieurs problèmes:

1- Levine ne fournit pas de critère pour dire à quel moment et à quelles conditions une expérience phénoménale est suffisamment proche (un analogue) de la nôtre pour qu'elle constitue un contre-exemple pertinent face au fonctionnalisme. Sans ce critère, le compromis de Levine reste une affaire d'intuition.

2- Ce compromis va à l'encontre de la position défendue par Levine lui-même, dans « *Materialism and Qualia: The Explanatory Gap.* » (1983), qui voulait montrer que cette contrainte de nécessité à travers tous les mondes possibles était la base d'un argument épistémologique contre le matérialisme. Ce problème pourrait être minimisé en disant que le compromis offert par Levine est une étape nécessaire, mais non suffisante pour défendre le fonctionnalisme (ou le matérialisme). Ainsi, le fonctionnalisme se doit dans un premier temps d'être capable d'expliquer les expériences analogues, mais ultimement, il devra expliquer toutes les expériences.

3- Ce compromis montre que Levine ne différencie pas les diverses thèses

fonctionnalistes. Ainsi, le fonctionnalisme de sens commun semble être obligé de d'expliquer *toutes* les expériences, car sa position « généreuse » face à la réalisation multiple l'oblige à dépasser largement le cadre de notre expérience « normale ». Minimale, il doit passer le test du compromis de Levine, mais ce n'est pas nécessairement le cas de toutes les thèses fonctionnalistes. Ainsi, certains fonctionnalistes empiriques penseraient que le compromis de Levine s'attaque à un faux problème car leur théorie de la mentalité n'a jamais cherché à expliquer les mentalités « analogues », mais seulement les mentalités « normales ». Comme Dick et Jane n'ont pas de cellule comparatrice dans le cas 2B, il n'est pas clair qu'ils soient les agents « normaux » que leur théorie cherche à expliquer. Cette réplique potentielle de la part du fonctionnaliste empirique est faible, car il n'est pas clair que des agents « normaux » existent (sauf en tant qu'entité statistique). Cette réplique illustre les limites de la critique de Levine. L'argument de celui-ci en général ne fait que montrer l'échec du fonctionnalisme du sens commun, car ce dernier est le seul qui soit contraint à accepter la nécessité kripkéenne, nécessité entre les mondes possibles, ou minimale, les expériences analogues. Ce n'est pas nécessairement le cas pour les autres sortes de fonctionnalisme qui ne cherchent pas à fournir une théorie générale de l'esprit, mais bien une théorie de la mentalité humaine « normale ». Comme Hardin n'est pas un fonctionnaliste du sens commun, car les fonctions sont fondées sur l'organisation physique du cerveau et non sur des rôles attribués par le sens commun, je ne suis pas certain que le compromis de Levine ou l'argument plus large de nécessité présenté dans son autre article s'appliquent à Hardin, puisque ce dernier a simplement à répliquer que sa prétention était beaucoup plus limitée que celle du fonctionnalisme du sens commun.

Levine était peut-être conscient des limites de son compromis ou simplement de la pertinence réelle de 2C car il concède qu'on pourrait peut-être comprendre 2C comme montrant seulement la plausibilité du résidu qualitatif dans 2B⁷¹. Reste le cas

⁷¹ « *But suppose one does not share my view about the legitimacy of 2C itself as a case of inverted qualia. We can still use 2C to buttress support for the coherence of 2B. Since case 2C shows that there is a separable feature of our visual qualia -the qualitative residue- in terms of which we can compare our qualia of both Dick's and Jane's, it seems possible that one could have an inversion with respect to just this feature but not with respect to the dimension of polarity. Thus, even if 2C*

2D, décrit ici par Levine, qui est le cas le plus audacieux jusqu'ici car il pousse encore plus loin l'inversion.

Dick's comparator cell works the way everyone else's does. It goes into state P when it detects increased activity and state N when it detects decreased activity. His r-g and y-b channel are inverted with respect to ours. So, reflectances from red and yellow objects cause decreased activity in the relevant channels, and vice versa for green and blue objects. Furthermore, and this is the crucial part of the case, all those stimuli across the various sensory modalities which in normal people cause those sensory organs to trigger the P-response in the comparator cells (...) cause Dick's sensory organs to trigger the N-response. Thus warmth itself, and other properties which in normal people trigger the P-response, trigger the N-response in Dick. (Levine, p.35)

Ce que Levine inverse ici, c'est la réaction produite par la cellule comparatrice. Ainsi, la cellule comparatrice continue à être dans un état *positif* lorsque le canal chromatique est *positif* (même si dans le cas de Dick l'inversé, le positif du canal chromatique est produit par des couleurs qui produisent chez l'observateur « normal » l'inhibition du même canal chromatique). Toutefois, la réaction produite par l'état *positif* de la cellule comparatrice est maintenant la réaction produite par l'état *négatif* de la cellule comparatrice chez l'observateur « normal ». La chaleur est donc associée à l'état négatif de la cellule comparatrice. Ainsi Dick dira que la tomate est rouge, mais son expérience qualitative est verte, et l'état négatif du canal chromatique est associé à l'état négatif de la cellule comparatrice qui, parce qu'elle aussi est inversée, sera associé à la chaleur. Dick dit que la tomate est rouge et chaude, même si son expérience qualitative des couleurs et ses réactions associées sont inversées. Levine croit que ce cas d'inversion détruit l'argument anti-inversion de Hardin. Hardin voulait tracer un lien « nécessaire » entre l'expérience qualitative et l'expérience du chaud/froid en montrant que la distinction chaud/froid était fondée sur un système physique. Ce que Levine veut montrer c'est que le lien entre l'état positif de la cellule comparatrice et la chaleur est tout aussi contingent que l'association de l'état positif du canal chromatique avec le rouge ou le jaune et qu'en déplaçant l'inversion (ou dans ce cas-ci en la dédoublant), nous aurons affaire à deux systèmes fonctionnellement isomorphes,

were not to count as a legitimate case of inverted qualia in its own right (which as yet to be demonstrated) it still helps to show that case 2B doesn't really verge on the « conceptually incoherent » after all. » (Levine, p.34)

mais qui sont phénoménalement distincts.

Comme Levine le souligne, Hardin pourrait répondre que nous ne sommes pas vraiment libres de dissocier l'état positif de la chaleur et que la contingence proposée par Levine n'est pas une possibilité réelle⁷². Par contre là où Levine marque un point, c'est quand il dit qu'il n'y a pas de relation nécessaire entre l'expérience qualitative de la chaleur et notre énonciation du mot « chaud ». L'opération du mécanisme d'association de l'expérience qualitative de la chaleur et de notre réaction verbale est tout aussi arbitraire que l'association de notre expérience qualitative des couleurs et de notre réaction verbale les concernant. Ainsi, Hardin ne peut pas accorder une primauté de vérité à l'expérience de la chaleur alors qu'il admet que l'erreur est possible à propos de notre expérience de la température. Autant c'est l'habitude qui fera dire à Dick que la tomate est rouge, même s'il la « voit » verte, de même, c'est l'habitude qui lui fera dire que la tomate lui donne une impression de chaleur, alors qu'il la « sent » froide.

Hardin voulait montrer que ce sont des considérations empiriques qui permettront de régler la question de l'inversion du spectre, mais les cas d'inversion proposés par Levine tendent à montrer que ce ne peut être le cas, car l'anti-fonctionnaliste pourra toujours répondre aux nouvelles asymétries de l'espace phénoménal en redécrivant les liens relationnels des qualia⁷³. Levine termine son argument en retournant à une critique myope du fonctionnalisme, en disant que même si les fonctionnalistes réussissaient à montrer que l'inversion n'était pas possible chez l'homme, on pourrait toujours imaginer une créature où le lien entre qualia et états fonctionnels est contingent. Je dis que cette critique est myope car elle assume encore une fois que *tous* les fonctionnalistes chercheront à expliquer des hypothétiques qualia martiens, alors que dans la réalité, plusieurs fonctionnalistes ne chercheront pas à élargir autant le domaine de la réalisation multiple. L'argument de

⁷² En fait il semble intuitivement difficile de défendre l'idée que la chaleur pourrait correspondre à une inhibition alors que dans tous les systèmes physiques elle correspond à une augmentation de l'énergie ou de « l'excitation ». Par contre le lien entre le rouge et la chaleur ne montre pas selon moi une telle nécessité intuitive.

⁷³ En fait, Levine pense que Hardin, en croyant que les faits empiriques régleront le problème, souffre du même manque d'imagination que ce dernier attribue aux anti-fonctionnalistes qui affirment que les cas d'inversion décisifs sont simples à construire.

Levine est efficace, mais seulement envers le fonctionnalisme de sens commun, ou toute thèse fonctionnaliste qui cherche à être généreux dans sa description des systèmes pouvant instancier notre type de mentalité. Un fonctionnaliste empirique qui définit les conditions de réalisation fonctionnelle de notre expérience sur de bases neurologiques humaines, ne cherchera pas à découvrir si les Martiens ont des qualia ou si la relation entre leurs qualia et leurs états fonctionnels est contingente, car nous n'aurons pas affaire au même système fonctionnel. Le commentaire de Levine perd donc de la force. Même s'il dit que « *so long as we can make sense of the notion that some creature could have a feature of experience that is both qualitative and independent of functional role, we break the conceptual tie between qualitative and functional role.* » (Levine, p.36), il n'a pas réussi à faire sens justement de cette créature car il n'a pas décrit un cas d'inversion assez fort pour nous faire croire que les asymétries phénoménales pouvaient être voilées par des inversions supérieures.

Même s'il réussissait à le faire, on pourrait alors imaginer une asymétrie supérieure, mais Levine pourrait sûrement jouer le même petit jeu et déplacer l'inversion toujours plus haut. La question fondamentale devient donc, jusqu'où Levine est-il prêt à aller? Il est fort douteux qu'il puisse impunément inverser l'espace qualitatif chromatique, l'espace qualitatif de la chaleur, etc., sans que cela ait le moindre effet sur le comportement, et même s'il réussissait à le faire, il ne ferait que montrer l'épiphénoménisme de l'expérience phénoménale, conclusion à laquelle il ne voudrait sûrement pas arriver. Levine ne peut pas d'une part affirmer qu'il pourra toujours trouver une inversion supérieure indétectable, et d'autre part affirmer l'efficacité causale comportementale des qualia.

4.1.5 Problèmes du matérialisme en général

Malgré les tensions présentes dans l'argument de Levine, ce dernier (p.36) a raison quand il affirme que, alors que la possibilité de l'inversion des qualia (chromatiques ou non) implique l'échec du fonctionnalisme, l'impossibilité de l'inversion ne constitue pas une preuve de l'efficacité du fonctionnalisme. Les cas de qualia absents en sont un flagrant contre-exemple. Levine revient ici à l'argument

kripkéen déjà examiné dans « *Materialism and Qualia: The Explanatory Gap* » (1983) et revient à sa conclusion que l'argument reste épistémologique. Je ne reviendrai pas ici à cet argument répété dans « *Cool Red* », car nous l'avons déjà examiné dans une section précédente. Il suffit de dire que le fonctionnaliste, même s'il réussit à résoudre tous les problèmes liés aux qualia, devra expliquer *pourquoi* notre activité est associée à une expérience phénoménale alors que l'on pourrait ne pas avoir une telle expérience et être fonctionnellement isomorphe à l'agent normal. Encore une fois, Levine semble se faire l'apôtre de l'inefficacité causale des qualia, ce qui semble étrange pour le pro-qualia qu'il se prétend être. Les téléofonctionnalistes qui cherchent à trouver la raison d'être des différents éléments fonctionnels (souvent en termes d'adaptation évolutive) diront que la conscience, ou l'expérience phénoménale, accorde un autre genre de connaissance de l'environnement. Ainsi, l'humain et l'hypothétique zombie ne seraient pas fonctionnellement isomorphes car la conscience de l'humain joue un rôle fonctionnel, un rôle causal.

Levine termine son article en traçant le plan de recherche (qu'il juge quasi impraticable) qui serait le plus satisfaisant à ses yeux: « *In my view what is needed is a good account of why, even if materialism is true, we are constitutionally incapable of seeing that [consciousness is reducible]. That is we need a good materialist explanation of the explanatory gap.* » (Levine, 1991, p.39). On pourrait bien sûr trouver plusieurs philosophes qui répondraient à Levine qu'il est bien évident que la conscience est réductible, mais cette réponse n'est pas satisfaisante. Il faudrait selon moi voir dans la suggestion de Levine, une visée programmatique pertinente qui pourrait être intégrée dans une description fonctionnaliste. À première vue, le téléofonctionnaliste est le mieux placé pour utiliser ce genre d'approche⁷⁴ car l'histoire de l'évolution pourrait peut-être fournir les outils descriptifs nécessaires pour expliquer cette incapacité cognitive de retour sur notre expérience. Toutefois, à ma connaissance, le travail reste à faire⁷⁵, mais il est plus prometteur que ne le sous-

⁷⁴ Comme je l'ai déjà dit dans l'introduction, une approche externaliste serait aussi intéressante, mais elle ne sera pas examinée dans ce mémoire.

⁷⁵ Levine mentionne que Rey dans un manuscrit inédit fournit une tentative en ce sens.

entend Levine. Dans la section 5 du mémoire, j'examinerai l'article programmatique de Sober qui tentera de bâtir un téléofonctionnalisme répondant aux défis lancés par Levine, mais, auparavant, voyons comment Hardin s'attaque plutôt à la cohérence de l'argument de Levine en tant que telle.

4.2 Critique de Hardin

4.2.1 Épiphénoménisme des qualia

Hardin nous offre dans le même numéro de *Philosophical Psychology*⁷⁶ une brève réponse à la critique de Levine. De manière grossière, Hardin essaie de faire deux choses. Il veut montrer certaines failles de la position dualiste et montrer que la description du problème offerte par Levine ne diminue pas la force de ce qui était proposé dans les faits par Hardin.

Hardin croit que le cas 2D proposé par Levine (soit de l'inversion des jugements de température associés à la cellule comparatrice) est un argument frappant en faveur de l'épiphénoménisme des qualia (en fait il précise que c'est un cas encore plus frappant que le cas qu'il avait lui-même proposé en 2A). Or Hardin croit que la thèse épiphénoméniste est paradoxale: « *the standard objection to epiphenomenalism has always been that it flies in the face of the very experiences that it was meant to accomodate* »(Hardin, 1991, p.41). En effet, il semble difficile de comprendre comment on pourrait concevoir des qualia épiphénoménaux alors qu'on semble obligé (si on accepte l'inversion du spectre) de nier leur efficacité causale. Si on nie leur efficacité causale, par quoi peut-on reconnaître notre objet d'étude? Le pro-qualia, Levine ici, en essayant de protéger la spécificité et l'ineffabilité de l'expérience phénoménale, se voit contraint de le retirer de la chaîne causale, mais en ce faisant, comment pouvons-nous identifier ces entités? Et *a fortiori* comment pouvons-nous dire que ces entités sont épiphénoménales? Les analogies sont hasardeuses, mais l'épiphénoméniste semble être dans la position de celui qui cherche à pointer du doigt un objet qu'il définit comme étant invisible. Or Levine, en défendant les cas d'inversion du spectre semble devoir adopter

⁷⁶ Hardin, C.L. « Reply to Levine » in *Philosophical Psychology*, 4(1) 1991.

l'épiphénoménisme et le paradoxe qui l'accompagne⁷⁷. Son cas 2D accentue l'inefficacité causale des qualia en dédoublant l'inversion. Hardin n'est pas en train d'avancer l'idée que les qualia n'existent pas, mais au contraire que l'expérience phénoménale joue un rôle causal et qu'une inversion indétectable (qui sonnerait le glas du fonctionnalisme) est donc impossible.

4.2.2 L'implausibilité de l'inversion de Levine

Les avocats de la possibilité d'une inversion du spectre indétectable sont donc pris au piège, mais ce n'est pas le seul problème qui devra les occuper. Hardin critique de manière plus large l'usage de l'expérience de pensée de l'inversion du spectre. Son opinion de ce genre d'argumentation est très clair: « *Now just as one must in science distinguish between real experiments and thought experiments, one must in philosophy distinguish between genuine and specious alternative possibilities.* »(Hardin, p.44). Hardin est peut-être méfiant des expériences de pensée en général, mais il se méfie surtout de l'extension de certaines expériences de pensée⁷⁸. De manière minimale on peut dire qu'il croit que l'expérience de pensée de Levine pose problème.

En fait l'argument de Levine se fait en deux temps. Il commence par essayer de montrer qu'une inversion indétectable est possible chez l'homme, ce qu'il ne réussit pas à montrer de manière concluante. Toutefois il se sert du doute introduit ici pour proposer que, même si nous réussissions à nier l'inversion chez l'homme, nous pouvons imaginer une créature où l'inversion serait possible. Hardin réplique en montrant que l'inversion n'est définitivement pas possible chez l'homme et qu'il n'est pas pertinent de parler d'autres créatures. Hardin concentre son attaque sur 2D

⁷⁷ Hardin définit très bien le problème en termes intentionnels: « *Indeed, if sensory qualities were epiphenomenal, since the qualities themselves could play no role at all in our coming to have beliefs about them. How they could then serve as contents of our beliefs becomes an impenetrable mystery.* »(Hardin, p.41)

⁷⁸ En fait il ne croit pas que l'on devrait se préoccuper de tous les mondes possibles: « *Indecent accounts are, of course, always available; a little tinkering ingenuity will always yield a logically inequivalent contender, but nobody is likely to pay it any serious heed. One of the most important differences between a bonafide theoretical alternative and one that is cooked up for the case at hand is that the former permits extensions to fresh data and conditions as well as connections with other theories, whereas the latter does not.* » (Hardin, p.44). Tout l'article cherchera à montrer que les expériences de pensée de Levine tombe dans la seconde catégorie et devraient donc être négligées.

qui est selon Levine l'argument le plus fort. Hardin avance que l'on ne peut pas inverser la réaction de chaleur et de froid sans aussi inverser les affects associés à la chaleur. De la même manière que Hardin voulait résoudre le problème de l'inversion du spectre en disant que l'expérience chromatique était associée à l'expérience de la chaleur, il veut ici répondre à Levine qui disait qu'on pourrait inverser l'expérience « thermique », en disant que l'expérience « thermique » est associée à d'autres émotions. De la même manière que le rouge semble être « chaud », le « chaud » semble être excitant. Or si Levine veut inverser nos associations de chaleur, il devra aussi inverser les émotions, mais comment affirmer que d'être excité ne modifie pas le comportement? Ceci semble être absurde et indique que Levine ne pourra pas déplacer l'inversion impunément⁷⁹.

Le problème de l'argument de Hardin est identique au problème de son argument concernant l'association de couleurs à certaines réactions face à la température. Comme Hardin le dit lui-même, tout ceci s'appuie sur certaines preuves empiriques, mais ceci ne pourrait être que circonstanciel. Il n'a pas montré de lien nécessaire entre une réaction à la chaleur et les émotions. Il y a tout autant d'indices que différentes cultures associent différentes couleurs à différents « états d'âme ». La question reste donc ouverte. Toutefois, la piste qu'il présente est intéressante car elle place la résolution du problème dans le domaine empirique, ce qui revalorise le matérialisme (du moins sous certaines formes).

Comme le souligne Hardin, une inversion radicale telle que celle proposée par Levine est peu plausible, car on imagine mal comment on pourrait tout inverser sans que cela affecte le comportement. Or une des prémisses de l'inversion du spectre est sa supposée plausibilité intuitive. Le problème s'aggrave avec la possibilité d'autres créatures, car on doit alors accepter la plausibilité d'une créature différente de nous, mais dont l'expérience phénoménale est analogue à la nôtre tout

⁷⁹ « There is, we must recall, some evidence that exposure to « warm » colors engenders an increase in various physiological indicators, whereas exposure to « cool » colors has the opposite effect (...). So an inversion of the requisite thoroughness, if it is to remain undetected, would require that those stimuli that generate depression in Dick bring about arousal in Jane, and vice-versa, and yet they behave in exactly the same way. Such a severing of affective states from behavior is surely epiphenomenalism with a vengeance -my case 2A redux. »(Hardin, p.46-47)

en étant détachée de tout rôle fonctionnel. Si nous trouvons difficile d'imaginer pour l'homme que l'expérience phénoménale est détachée de tout rôle causal, comment défendre l'idée qu'une créature différente de nous a une expérience phénoménale analogue à la nôtre et que cette expérience ne joue aucun rôle causal? Et c'est ici que je ne suis pas certain de la position de Hardin face aux expériences de pensée. Il semble s'en méfier car on pourrait toujours en appeler à un cas supérieur se rapprochant toujours un peu plus d'un scepticisme généralisé⁸⁰. C'est le même problème que celui soulevé lors du débat autour de « *Materialism and Qualia: The Explanatory Gap* ». On ne peut pas refuser *a priori* des résultats comme ceux présentés par Hardin à propos de la relation entre l'expérience des couleurs et celle de la chaleur, tout en maintenant que d'autres résultats « scientifiques » ont un statut différent. La question trouvera sa réponse de manière empirique et les cas alternatifs soulevés par Levine sont donc non pertinents. •

4.2.3 Résidu qualitatif

Hardin en revient aussi au fameux « résidu qualitatif ». Contrairement à l'intuition, Hardin concède la possibilité d'un certain résidu qualitatif, mais réduit la spécificité de celui-ci. En effet, Hardin continue d'affirmer qu'il est difficile de concevoir un résidu qualitatif du vert après que l'on ait inversé son expérience phénoménale chromatique, « thermique » et affective. Que resterait-il de spécifique à l'expérience pour nous permettre de dire que nous avons encore affaire à un quale de vert? En fait Hardin se demande ce qui permet à Levine d'affirmer que ce « résidu » hypothétique est différent pour le vert et le rouge. L'avenue empruntée par Hardin est étrange (et ne serait probablement pas acceptée par Levine car elle réduirait trop la spécificité de l'aspect qualitatif); plutôt que de simplement rejeter la possibilité de résidu qualitatif (ce qui serait cohérent avec son physicalisme), il en fait une espèce de propriété généralisée et unique. Ainsi toutes les sensations pourraient être accompagnées d'un « résidu qualitatif », mais du même « résidu qualitatif » (et non

⁸⁰ « *Surely this sort of answer, delivered in its full particularity, would deliver some epistemic relief. It would not rule out all alternative explanations, but that is quite beyond the capacity of any scientific account; one could always have recourse to evil demons, a Berkeleian God, or brains-in-vats.* »(Hardin, p.48)

un résidu qui varierait d'une expérience à une autre). Hardin ne peut pas prouver l'unicité de ce résidu, mais ce n'est pas son but. Son but est de montrer que la même incapacité introspective qui donne tout son sens à l'argument de l'inversion, empêche aussi de Levine de prouver que chaque couleur a un « résidu » particulier qui permettrait de la distinguer des autres couleurs sans utiliser de termes neurologiques ou fonctionnels. Hardin concède qu'il y a peut-être un aspect qualitatif en marge de l'explication fonctionnaliste, mais cette exclusion le rend tout aussi obscure pour le pro-qualia ce qui fait des arguments de ce dernier des hypothèses non-vérifiables.

Ceci illustre le paradoxe de ce genre d'argument pro-qualia car les pro-qualia fondent la plausibilité de leur argument sur leur valeur intuitive, mais arrivent à des conclusions totalement contre-intuitives. Par contre Hardin croit qu'il faut rendre compte de l'intuition que « *there is nothing it is like to be that creature* ». Aucune explication matérialiste ne pourra pleinement rendre compte de ce que c'est réellement être une chauve-souris. Hardin croit que cela reste un problème pertinent, mais plutôt que d'y voir l'échec du matérialisme, il faut plutôt voir comment le matérialisme (et dans le cas de Hardin, le physicalisme) peut rendre les qualia tellement génériques qu'ils constitueront une dimension unique partagée de tous les objets ou expériences. Ici Hardin offre une brève description d'un panpsychisme qui ferait des qualia (ou du « mental » en général) l'envers (le « *flip-side* ») de toute matière.

4.2.4 Conclusion

Je ne commenterai pas le panpsychisme de Hardin. L'argument (s'il y en a un) n'est pas assez étoffé pour qu'on en examine la pertinence ici. En fait, je crois que nous devons nous en tenir à la conclusion modeste selon laquelle Levine ne peut prouver qu'un quale peut rester « vert » (et non un résidu plus générique) après avoir inversé toutes ses propriétés. À mesure que l'on repousse l'inversion toujours plus profondément, la spécificité d'un hypothétique résidu qualitatif devient de plus en plus chimérique et sans cette spécificité phénoménale, il n'y a pas d'inversion.

On pourrait croire que Hardin ne fait que louvoyer entre les critiques de

Levine, mais ce n'est pas le cas. L'argument de Hardin n'est insatisfaisant que si on le considère comme un fonctionnalisme de sens commun, ce que ne fait pas Hardin. En fait, le commentaire initial de Hardin sur l'inversion du spectre ne se voulait pas une défense du fonctionnalisme, mais un indice de l'efficacité du physicalisme à résoudre le problème de l'inversion du spectre.

Comme je l'ai déjà mentionné, Hardin ne se préoccupe pas vraiment de savoir si sa thèse est fonctionnaliste ou physicaliste, mais il souligne les lacunes du fonctionnalisme parce que ce dernier offre des explications trop abstraites qui pourraient trouver des instanciations farfelues⁸¹. Toutefois, et ici Hardin souffre du même manque de précision que Levine, ce ne sont pas tous les fonctionnalismes qui cherchent à obtenir ce degré d'abstraction. La description qu'offre Levine de son objectif pourrait fort bien être accommodée dans une explication téléofonctionnaliste:

I do not much care whether my little story is taken to be a functional or a physiological account. One reason for labelling it as the latter rather than the former is to suggest that I conceive of it as embedded in an overall design for a brain that is in turn attached to an organism that must cope with an environment. (Hardin, p.45)

Le téléofonctionnalisme cherche aussi à faire sens de certains aspects de la mentalité en termes d'adaptation. L'objectif minimal de Hardin est de montrer comment on peut rendre compte physiquement de la *différence* entre voir vert et voir rouge. Il ne cherche pas à montrer que sa théorie peut épuiser ce que c'est de voir rouge plutôt que vert, mais s'il a raison et qu'il peut montrer qu'il y a une différence nécessaire entre voir vert et voir rouge, alors il n'y a pas d'inversion indétectable. Si cela fonctionne, alors l'inversion du spectre ne constitue pas un contre-argument au fonctionnalisme. Selon Hardin l'avantage de son physicalisme versus la position pro-qualia, c'est que la neurologie donne au moins les conditions nécessaires à la perception de la couleur, alors que le pro-qualia peut difficilement affirmer qu'il

⁸¹Hardin se méfie du fonctionnalisme en général : « *one of the reasons that functional accounts of sensory events have often appeared implausible is not necessarily that they were functional accounts as such, but because the functions were described at a level sufficiently high and abstract as to permit absurd instantiations.* » (Hardin p.42) Toutefois il semble voir que son genre d'explication pourrait aussi être comprise comme une thèse fonctionnaliste: « *as Levine rightly points out, a physiological explanation is a functional explanation of a particular detailed type.* » (ibid).

existe des qualia de couleurs sans réseaux neuronaux.

Le danger de ce genre de pari c'est qu'il est attaquable *a priori* et *a posteriori*. Ainsi, malgré son enthousiasme, Hardin convient que l'expérience phénoménale restera peut-être un mystère, un genre de point aveugle épistémologique. De plus, en se plaçant fermement dans le camp empirique, toute sa théorie est la victime potentielle de résultats empiriques futurs qui viendraient contredire ses résultats. Que doit-on retenir de l'argument de Hardin? Cela dépend, me semble-t-il, de sa position initiale! Le matérialiste dira que l'inversion se révèle être un faux problème, alors que le dualiste dira que le fonctionnalisme reste incapable de rendre compte du « résidu qualitatif » (quel qu'il soit). Mais de manière provisoire, je crois que le constat que l'inscrutabilité des qualia tient autant pour le matérialiste que pour le dualiste, et reste un point convaincant; le fonctionnaliste, sans pour autant pouvoir crier échec et mat, peut au moins murmurer échec et pat.

5 Téléofunctionnalisme

4.3 Argument de Sober

5.1.1 Introduction

À travers les différents arguments examinés jusqu'ici, une constante s'est dégagée: le fonctionnalisme du sens commun semble être condamné à l'échec dans la mesure où il est incapable de restreindre son domaine d'application. Ce manque de précision le rend sujet à une pléthore d'attaques se basant sur la nécessité d'une distinction fine entre les états mentaux. Le téléofunctionnalisme se présente comme une alternative pertinente à ce fonctionnalisme trop « libéral » sans tomber dans une caractérisation « chauviniste ». Elliot Sober, dans « *Putting the Function Back into Functionalism* »⁸² offre une description optimiste (ou au moins programmatique) de ce téléofunctionnalisme qu'il juge supérieur au fonctionnalisme « classique » parce que celui-ci donne plus de contraintes à la notion de fonction. Cet ajout de contraintes, selon Sober, permettrait au téléofunctionnalisme d'éviter les écueils sur lesquels échouait par le passé le fonctionnalisme. Bien que cet article de Sober n'offre pas la description la plus détaillée du téléofunctionnalisme⁸³, elle offre au moins de manière synthétique une réponse aux défis lancés par Levine au matérialisme.

Le point de vue de Sober est clair: « *Functionalism got off on the wrong foot.* » (Sober, p.63). Le concept « classique » de fonction n'a pas été bien circonscrit. Contrairement à la majorité des commentateurs, la cible de sa critique n'est pas tant le fonctionnalisme de sens commun, mais ce que Braddon-Mitchell et Jackson appelait le *machine functionalism* et que Sober appelle le *Turing Machine Functionalism*. Étrangement, Sober ne parle pas des autres sortes de fonctionnalisme comme si toutes les variantes que l'on pouvait identifier étaient subsumées sous la notion de fonctionnalisme machine. Cette caractérisation n'est pas fautive en soi.

⁸² Extrait de « *Panglossian Functionalism and the Philosophy of Mind* » in *Synthese*, volume 64 (2), 1985. Paru ensuite dans « *Mind and Cognition* » ed. W.G. Lycan 1^{er} éd. (1990) et 2^e éd. (1999). Le texte étudié ici est celui de la seconde édition de Lycan.

⁸³ Pour un examen détaillé du concept de fonction en termes téléologiques, on peut se référer entre autres au livre « *Consciousness* » de Lycan (1987).

Dans la mesure où le fonctionnalisme machine est considéré comme appartenant au fonctionnalisme empirique tout en ayant un degré d'abstraction plus élevé que les autres fonctionnalismes empiriques, il semble être un compromis adéquat entre le fonctionnalisme analytique et le fonctionnalisme empirique. Braddon-Mitchell et Jackson voyait dans ce « compromis » une tension, voir même une incohérence, intrinsèque au fonctionnalisme machine, d'où la nécessité de son rejet⁸⁴.

5.1.2 *Fonction biologique plutôt que mathématique*

Sober ne se préoccupe pas tant de ces précisions taxinomiques. Ce qui l'intéresse c'est le modèle sur lequel on base notre compréhension du concept de fonction. Il croit que l'échec du fonctionnalisme « classique » (qui je le rappelle correspond chez Sober au fonctionnalisme machine et non au fonctionnalisme de sens commun) vient du fait qu'il emprunte le concept de fonction aux mathématiques plutôt qu'à la biologie. Cette erreur aurait selon Sober une source historique. Le fonctionnalisme serait apparu suite à deux intuitions. L'intuition négative est que les propriétés psychologiques ne sont pas identiques type-type à des propriétés physiques. Cette intuition est liée au credo fonctionnaliste de la réalisation multiple. La seconde intuition, positive ici, est que l'on peut caractériser une propriété mentale selon le rôle causal qu'elle joue. Jusqu'ici Sober croit que le fonctionnalisme ne se trompe pas. Là où il a commencé à errer, c'est lorsqu'il a joint ces deux intuitions à l'idée séduisante que l'esprit est une sorte d'ordinateur. Les capacités cognitives sont alors identifiées à des « programmes » d'ordinateurs. Il suffit alors de décrire les programmes en tant qu'instanciation de table machine et nous pouvons alors comprendre les computations de l'esprit. Si deux « systèmes » instancient la même table, ils ont le même état mental. Ceci correspond donc à l'intuition négative (les tables peuvent être instanciées dans divers systèmes) et à l'intuition positive (on identifie les bonnes tables selon leur rôle causal).

Le problème est qu'en considérant la fonction en terme de programme, on la

⁸⁴ Cette incohérence venait du fait que, en tant que fonctionnalisme empirique, le fonctionnalisme machine s'intéresse au substrat physique qui permettra de réaliser le « *software* », la table machine qu'il cherchera à établir, mais en essayant de s'éloigner du problème du chauvinisme, il définit le « *software* » comme étant relativement libre face à la réalisation physique particulière. Ceci rend le

caractérise en termes mathématiques: tout ce qu'il faut pour que l'on puisse dire que deux systèmes ont les mêmes états mentaux est un isomorphisme abstrait de leurs programmes. Or la connexion entre ces programmes et les états mentaux est loin d'être nécessaire comme le montre divers arguments anti-fonctionnalistes. Comme Block et Fodor l'ont montré à plusieurs reprises⁸⁵, de la même manière que l'on peut facilement imaginer que deux êtres soient dans le même état mental tout en étant dans différents états physiques, on peut imaginer que deux êtres soient dans le même état mental tout en n'étant pas dans le même état computationnel. Sober croit que ce genre d'attaque est beaucoup plus décisif pour le fonctionnalisme machine qu'il ne l'est pour le téléofonctionnalisme.

L'avantage du téléofonctionnalisme est qu'en empruntant le concept de fonction à la biologie plutôt qu'aux mathématiques, il est en mesure d'identifier les rôles causaux *pertinents*, d'ajouter des contraintes qui permettront au téléofonctionnalisme de répondre aux arguments standards, tout en restant assez abstrait pour maintenir la prémisse de la réalisation multiple. Dans le reste de son article, Sober explore différentes manières de comprendre la « vraie » notion de fonction.

Comme le dit Sober (p.64) l'isomorphie téléofonctionnelle n'a rien à voir avec l'identité entre deux programmes⁸⁶ car elle n'est pas basée sur une isomorphie entre structures abstraites, mais sur la notion de « raison d'être ». Secondement, pour que le fonctionnalisme soit une thèse intéressante, il ne doit pas seulement fournir une explication de la mentalité, mais de tout l'organisme⁸⁷, ce que les autres thèses fonctionnalistes négligeaient ou, minimalement, prenaient pour acquis. Sober prend l'exemple du système digestif. Si la réalisation multiple tient, on doit accepter que la digestion se retrouve chez des créatures qui diffèrent largement de nous

fonctionnalisme machine trop libéral.

⁸⁵ voir Block et Fodor (1972) et Block (1978).

⁸⁶ Je suis conscient que le mot « isomorphie » est un terme technique en mathématique, mais il n'est utilisé ici que dans un sens large de « *sameness* ». Comme dans tout mon mémoire isomorphie veut tout simplement dire « ayant la même forme » et non un sens mathématique plus précis. Cette précision est pertinente pour tout le mémoire, mais elle est particulièrement importante concernant l'examen de Sober car ce dernier s'oppose à la caractérisation mathématique de fonction.

⁸⁷ « *Multiple realizability is just as much a property of digestion as it is of perception and*

physiquement. Le fonctionnalisme classique ne rejette pas cette idée *a priori*, mais là où la théorie de Sober se distingue, c'est qu'il croit que l'on peut attribuer téléofonctionnellement la digestion à des espèces qui diffèrent de nous, non seulement physiquement, mais aussi par leur structure fonctionnelle abstraite si on comprend cette fonction en termes de « programmes ». « *For a system to engage in the process of digestion is for it to engage in a process that has a certain (teleological) function.* » (Sober, pp.64-65).

Ce n'est pas la chaîne causale dans laquelle le système digestif en question s'inscrit qui importe (ce qui était le cas pour les autres sortes de fonctionnalisme) car on pourrait tracer une autre chaîne causale tout aussi cohérente, mais triviale. Ce qu'il faudra trouver, c'est le « vrai » rôle du système digestif, et ce « vrai » rôle trouve sa raison d'être en terme évolutif. Sober ne va pas encore jusque là, mais c'est l'objectif vers lequel il se dirige. L'exemple de la digestion devrait éclairer des problèmes comme celui de la perception. Ainsi la plupart des fonctionnalistes décrivent le problème de la perception en assignant les input en termes de longueurs d'onde, les output en terme de construction de représentation ou comportement, mais ce genre de description est plus réductionniste que fonctionnaliste.

Le fonctionnaliste devrait fournir une description plus abstraite, plus détachée des conditions de réalisation physiques particulières, s'il tient vraiment à la thèse de la réalisation multiple. On en revient au dilemme de l'assignation des fonctions. Si on définit la fonction de manière trop large, elle devient triviale, pouvant être assignée à presque tous les systèmes, mais si on la définit de manière trop stricte, on imagine difficilement comment on ne pourra pas être chauvin quant à l'esprit. Une description trop stricte fera de la mentalité un apanage exclusivement humain, conclusion peu séduisante. Alors comment le téléofonctionnalisme pourra éviter ces deux écueils? Sober espère que « sa » fonction sera assez abstraite pour permettre la réalisation multiple sans devenir triviale. La digestion qui offre, semble-t-il, un cas plus simple que celui de la perception offre une avenue de réponse. « *Roughly speaking, it seems plausible to say that all digestion involves extracting*

energy from the environment. »(p.66)⁸⁸. Ainsi Sober croit que la perception pourrait de manière analogue être comprise comme étant faite pour extraire de l'information de l'environnement. Ce genre de caractérisation place sous un autre jour la notion d'input dans le processus de la perception. Ainsi, il est inintéressant de dire que les input dans la vision sont des quanta d'énergie frappant la rétine, alors que ce qui devrait nous intéresser c'est que nous avons des input ayant des « *information-bearing properties* »(p.66). La question des longueurs d'ondes, des photo-pigments et des neurones ne devraient donc nullement intéresser la description (téléo)fonctionnaliste car elle ne concerne que l'une des réalisations physiques possibles du système.

5.1.3 Le téléofonctionnalisme et les descriptions *a priori*

Est-ce que cela revient à dire que le téléofonctionnalisme n'est pas un fonctionnalisme empirique? Bien que Sober ne se prononce pas précisément sur la question, il me semble que la réponse doit être négative. Ainsi, en décrivant la fonction en tant que relation à l'environnement (du moins dans le cas de la perception et de la digestion), une description purement «aprioristique» ne semble pas intéressante car il sera nécessaire, lors de la définition de la fonction en question, de comprendre dans quel environnement le système s'inscrit, ce qui est une tâche empirique. Toutefois, la description téléofonctionnelle reste assez abstraite pour permettre un degré intéressant de réalisation multiple.

Ce qui motive Sober est en fait l'intuition que l'on ne peut pas trouver des raisons *a priori* à cause desquelles une description de ce genre devrait être impossible. En fait la réponse de Sober aux arguments habituels est intéressante car elle décrit ces arguments comme étant en fait trop *a posteriori*. Comme il le souligne lui-même (p.66) peut-être que la « perception » se révélera une catégorie trop large pour être pertinente. Ainsi il ne serait pas pertinent de critiquer la psychologie fonctionnaliste en disant qu'elle ne fournit pas d'explication fonctionnelle de la

⁸⁸ Ceci n'est probablement pas suffisant pour définir la fonction en question (même si Sober n'offre rien de plus comme définition), mais ce genre de description nous donne une idée de ce que Sober entend par le *telos* d'une structure.

perception⁸⁹. La portée de l'argument de Sober est donc réduite. Comme il le dit lui-même à plusieurs reprises dans son article, il ne cherche pas ici à fournir une définition de ce qu'est une fonction mentale. Tout ce qu'il veut montrer est que le téléofonctionnalisme n'est pas incohérent, car les objections classiques faites au fonctionnalisme s'attaque peut-être en fait à un faux problème ou du moins à un problème auquel le téléofonctionnalisme ne cherche pas à répondre.

Le problème le plus épineux reste celui des qualia. Comme Block, il accepte une position réaliste face aux qualia. Toutefois, Sober croit que l'on peut être réaliste quant aux qualia tout en maintenant une thèse fonctionnaliste. Son argument ressemble, du moins au début, plus à une pétition de principes:

Your qualia and mine differ, even though they appear to play the same causal roles. Does it follow that Functionalism cannot capture the difference between greenish qualia in me and reddish qualia in you? No, this does not follow, for it may be the case that there is lots of structure lurking within that distinguishes the two. Perhaps we experience qualia as simples, but that does not mean they are simples. For all we know there may be lots of complexity that a functionalist theory may identify and thereby make out a difference between your qualia and mine. For this reason, I do not think we can conclude that Functionalism cannot handle qualia. (Sober, p.67)

Jusqu'ici, cette position pourrait concorder avec le genre d'argument fourni par Hardin à propos des réactions associées à la perception particulière de certaines couleurs. Toutefois, un tel parallèle est démenti lorsque Sober endosse l'idée que des états psychologiquement différents pourraient avoir une fonction *psychologique* identique (p.67). En d'autres mots, le comportement ne permet pas de distinguer les états mentaux. On ne doit donc pas comprendre le fonctionnalisme de Sober comme appartenant à la même famille que les autres thèses fonctionnalistes. Les autres thèses fonctionnalistes ont un penchant psychologique (surtout le fonctionnalisme de sens commun, mais cela vaut aussi pour les thèses fonctionnalistes parce qu'elles s'attaquent de manière privilégiée au problème *corps-esprit*). Toutefois l'approche de Sober se veut plus large. En se plaçant d'un point de vue biologique plutôt que psychologique, certaines incompréhensions le dérangent moins que pour le

⁸⁹ Comme Hardin l'avait fait auparavant, Sober préférerait sûrement une explication cherchant à donner sens à la perception des couleurs en particulier, pour ensuite s'attaquer à la perception du mouvement, etc. Tout comme Hardin, Sober semble privilégier une méthode particulariste, la division d'un faux problème fonctionnel en plusieurs petits problèmes qui trouveront tous une réponse

fonctionnalisme classique. Comme il le dit lui-même (p.68), il trouverait surprenant que toute différence psychologique, dans le cas de l'inversion du spectre une différence de qualia, se traduise par une différence fonctionnelle.

Le point de vue biologique accepte l'idée que certains éléments d'une organisation fonctionnelle soient des artefacts évolutifs ne jouant aucun rôle causal pertinent ou fonctionnel. Si ces artefacts restent intéressants pour la biologie c'est qu'ils sont des conséquences du système fonctionnel, ou qu'ils ont déjà été des conséquences du système. Le fonctionnalisme « classique » est donc une théorie idéale en tant qu'elle cherche à attribuer un rôle fonctionnel à *tous* les éléments du système, mais une telle perfection ne se retrouve nulle part ailleurs dans la nature. Ce genre de fonctionnalisme psychologique est décrit par Sober comme étant panglossien⁹⁰. Sober souligne que l'on ne pourra pas trouver une fonction à tous les éléments. Certains éléments peuvent donc apparaître fonctionnellement isomorphes, mais c'est seulement parce qu'ils n'ont pas de fonction réelle.

En fait, ce qu'il faudra faire sera une relativisation des fonctions. Ainsi, si le quale du rouge et le quale du vert semblent jouer le même rôle fonctionnel, c'est peut-être qu'ils sont isomorphes seulement par rapport au système psychologique étudié, mais leur rôle téléofonctionnel est peut-être différent. Ainsi, Dick, l'agent inversé, et Jane, sont peut-être fonctionnellement identiques si on ne fait qu'examiner leur comportement (et si on rejette le genre d'explication naturaliste de Hardin), mais ils seraient téléofonctionnellement distincts, en tant qu'ils n'extraient pas *le même genre d'information* de l'environnement. Dick et Jane ne sont pas téléofonctionnellement isomorphes car ils n'ont pas le même rapport face au monde. Toutefois, on devra tout de même fournir une explication de la raison d'être de cette distinction. Ce n'est pas exactement la position de Sober, mais elle semble être cohérente avec son téléofonctionnalisme: il dit que ce n'est pas tout qui aura une fonction, mais il ne montre pas que les qualia n'ont *nécessairement* pas de fonction. Tout ce que Sober nous demande de faire c'est de s'éloigner de la contrainte

téléofonctionnelle.

⁹⁰ En l'honneur du docteur Pangloss dans le *Candide* de Voltaire qui disait que tout avait une fonction qu'il cherche à optimiser.

panglossienne du fonctionnalisme classique. On peut ultimement assigner une fonction à tout, mais ce ne sont pas toujours des fonctions intéressantes⁹¹. Le fonctionnalisme classique cherche pourtant à trouver une fonction à tous les éléments. Comme il le dit lui-même, un tel idéalisme est utile lors du développement d'une théorie, mais il se révèle toujours être extrémiste et erroné lorsque la théorie mûrit. Le (téléo)fonctionnalisme en admettant qu'il ne peut pas fournir une fonction à tout mais qu'il peut expliquer ce point aveugle, montre ce genre de maturité.

5.1.4 *Épiphénoménisme des qualia*

Je ne vois pas comment la position de Sober se distingue d'une position épiphénoméniste envers les qualia. Toutefois, sa position reste intéressante car elle élargit le fonctionnalisme de manière à ce que l'esprit ne soit plus considéré comme un objet unique et spécial, mais comme un élément d'une explication plus large. Avec ce recul, les artefacts ne sont pas seulement des anomalies, mais des objets faisant nécessairement partie de l'évolution de n'importe quel système. Les qualia perdent donc leur spécificité en tant que problème insoluble, ils sont peut-être tout aussi « inutiles » (sans fonction identifiable) que la deuxième corne des rhinocéros javanais (exemple offert par Sober p.68). Ce n'est pas parce que le (téléo)fonctionnalisme ne trouve pas de fonction aux qualia, que l'on doit en conclure que cette théorie est fautive. En fait, là où la théorie de Sober semble être plus utile que l'épiphénoménisme, c'est qu'elle fournit la structure permettant d'expliquer *pourquoi* nous avons ces épiphénomènes en montrant qu'ils ont déjà joué un rôle causal pertinent. De plus, elle fournit le genre de cadre matérialiste envisagé par Levine qui demandait une explication matérialiste de notre incapacité d'avoir un accès objectif à nos qualia. Si le téléofonctionnalisme réussit à montrer pourquoi les qualia sont « vestigiaux », il fournit une explication de

⁹¹ Sober (p.68) utilise l'exemple (emprunté à Cummins) du cœur qui, parce qu'il a une certaine masse, contribue à la masse totale du corps. Toutefois, nous ne jugeons pas nécessaire de dire que voilà une fonction du cœur. Peut-être que l'on montrera que les qualia jouent un rôle fonctionnel dans la perception, mais que ce rôle est trivial et que la vraie raison d'être des qualia est ailleurs ou tout simplement absente. Le téléofonctionnalisme, en intégrant l'analogie de la théorie de l'évolution, accepte l'idée que ce ne seront pas tous les éléments du système qui auront une fonction *pertinente* (mais en tant que la simple existence d'un élément aura des effets particuliers sur le système, il est certain que l'on pourra lui trouver une fonction, toutefois cette fonction est peut-être triviale).

l'épiphénoménisme actuel des qualia.

La théorie de l'évolution sur laquelle se basera le téléofonctionnalisme fournira le cadre conceptuel permettant d'expliquer de manière satisfaisante certains mystères mentaux d'aujourd'hui, en les inscrivant dans une histoire plus large. Toutefois, Sober n'en est pas au point où il peut tracer le pont entre la biologie et le fonctionnalisme. Le lien est à ce point simplement analogique. Le concept de fonction s'inspirera du concept de fonction en biologie, mais le lien reste mince. Sober à la fin de l'article et dans d'autres écrits fournit un examen des théories biologiques et s'interroge sur les limites de concepts comme le « *fitness* ». Je n'examinerai pas cet élément ici, mais il suffit ici d'un avertissement de prudence: le parallèle avec la biologie ne doit pas induire le lecteur dans l'erreur de faire la réduction type-type des concepts de la biologie ou du téléofonctionnalisme à des propriétés physiques. Sober croit justement que l'intérêt du fonctionnalisme est de s'éloigner de ce genre de théorie de l'identité. Ce serait la seule manière, selon lui de garantir la réalisation multiple.

Le téléofonctionnalisme doit donc éviter le désir du fonctionnalisme habituel de fournir une fonction à tous les éléments de la structure, ce qui correspond à une explication idéale impossible dans les faits, et le désir d'inscrire physiquement les fonctions ce qui empêchera la réalisation multiple du système. Toutefois, le téléofonctionnalisme se devra d'être plus naturaliste que Sober le décrit car en cherchant la raison d'être d'un système, et en fondant cette raison d'être dans une relation avec l'environnement, son explication sera *nécessairement* liée à des contraintes empiriques, car l'environnement ne peut être exploré qu'a posteriori. Il n'est donc pas clair que le téléofonctionnalisme peut vraiment être non-réductionniste (du moins partiellement).

5.1.5 Avantages du téléofonctionnalisme

Le téléofonctionnalisme offre une réponse intéressante aux critiques anti-fonctionnalistes, car il offre un cadre théorique permettant d'accommoder les défis lancés par Levine au matérialiste. Ainsi, le téléofonctionnalisme décrivant la fonction en termes de « raison d'être » et de rôle plutôt qu'en termes causaux, c'est

une théorie qui pourrait éventuellement répondre à la question de Levine: *pourquoi* nos perceptions sont accompagnées de qualia⁹²?

De plus, parce que le téléofonctionnalisme accepte l'idée que ce ne sera pas tout qui aura ultimement une fonction pertinente, une « vraie » fonction, il peut intégrer un certain degré d'épiphénoménisme ou du moins il peut accepter qu'il ne pourra pas expliquer fonctionnellement (« computationnellement ») tout système; ceci a l'avantage de prémunir le téléofonctionnalisme contre la plupart des arguments classiques anti-fonctionnalistes. Cette acceptation du fait que certains éléments soient « vestigiaux » ou du moins n'aient pas de fonction, est selon moi le point le plus intéressant du téléofonctionnalisme, car il répond au défi de Levine: « *In my view what is needed is a good account of why, even if materialism is true, we are constitutionally incapable of seeing that [consciousness is reducible]. That is we need a good materialist explanation of the explanatory gap.* » (Levine, 1988, p.39). En montrant pourquoi la conscience ou les qualia sont « vestigiaux »⁹³, le téléofonctionnalisme pourrait répondre à Levine et ainsi, répondre à sa critique anti-matérialiste, comblant ainsi le gouffre explicatif.

Enfin, en définissant la fonction en termes de rapport à l'environnement, le téléofonctionnalisme est un questionnement qui se place hors du débat à propos des autres mondes possibles. Le téléofonctionnalisme n'a pas la prétention de rendre compte de toutes les fonctions dans tous les environnements possibles. Seul notre environnement est sujet d'étude possible pour nous. Des arguments comme l'inversion du spectre se révèle donc sous un autre jour, car la question devient alors de savoir si ces cas sont possibles dans *notre* monde, dans notre environnement. Cette question se réglera de manière empirique comme l'espéraient les philosophes naturalistes comme Hardin et non par des nécessités kripkéennes.

⁹² Tout ce que le fonctionnalisme classique pouvait faire c'était d'espérer définir les qualia en termes fonctionnels (ou éliminer les qualia), mais il ne pouvait dire pourquoi nous avons ces qualia plutôt que d'être de simples zombies. Mais le téléofonctionnalisme semble pouvoir répondre à la question de Levine: « *However, there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is why should pain feel the way it does!* » (Levine, 1983, p.357).

⁹³ Bien sûr le téléofonctionnalisme n'est pas lié à cette conclusion. Comme j'ai essayé de le faire plus tôt, le téléofonctionnaliste pourrait peut-être trouver une « téléofonction » à la conscience ou au

6 Conclusion

Comme nous avons pu le voir tout au long de ce mémoire, l'inversion du spectre est une épine douloureuse pour le fonctionnaliste. La solution facile serait de tout simplement rejeter la possibilité de ce cas d'inversion. Toutefois, un simple rejet de la possibilité *a posteriori* (ou possibilité réelle d'une inversion) ne semble pas satisfaisante. D'une part, ce rejet se traduit la plupart du temps par un rejet de l'inversion chez l'être humain basé sur une certaine compréhension du système nerveux central humain. Or comme l'ont dit la plupart des commentateurs examinés ici, ce rejet n'est pas suffisant pour se débarrasser de l'inversion du spectre; si le fonctionnalisme veut rester un cadre explicatif intéressant, il doit avoir les outils pour expliquer les mentalités non-humaines, même si ces mentalités sont à ce stade-ci hypothétiques. De plus ces mentalités hypothétiques doivent rester explicables par le fonctionnalisme s'il veut maintenir la réalisation multiple. Il ne peut donc pas rejeter le problème de l'inversion en général, en disant qu'elle est impossible chez l'être humain en particulier. D'autre part, il n'est pas clair que l'impossibilité *a posteriori* soit suffisante pour protéger le fonctionnalisme car l'argument anti-fonctionnaliste habituel se sert de la possibilité *a priori* de l'inversion du spectre, argument que des résultats scientifiques ne peuvent pas aborder.

Les anti-fonctionnalistes exigent donc du fonctionnaliste la preuve que l'inversion est impossible *a priori*, objectif inaccessible selon Levine. Que peut donc faire le fonctionnaliste ? S'il ne peut pas rejeter la possibilité de l'inversion, il doit montrer que la possibilité d'une inversion n'est pas suffisante pour détruire le fonctionnalisme. Shoemaker va en ce sens, mais sa description du problème est moins une défense du fonctionnalisme qu'un argument pro-physicaliste. En identifiant un état fonctionnel à une disjonction d'états physiques qui, eux, correspondent à des qualia, la détection de l'inversion passe par l'identification d'états physiques et non par des états fonctionnels. On peut donc se demander à quoi sert le fonctionnalisme, si ce n'est qu'un genre de vulgarisation de phénomènes plus complexes. Il est intéressant de noter que l'argument de Shoemaker avait déjà été

envisagé par Levine qui voyait en l'argument de l'inversion, un argument fort contre le fonctionnalisme, mais moins efficace contre le physicalisme. Toutefois, Levine conclut que le physicalisme n'est pas plus en mesure de combler le gouffre explicatif. Il serait donc intéressant de voir en quoi la thèse fonctionnaliste est plus intéressante que le physicalisme, si ce dernier est capable de rendre compte de l'inversion alors que le fonctionnalisme ne semble pas en mesure de le faire. La réponse est la réalisation multiple. Le physicalisme ne peut pas intégrer la réalisation multiple de manière aussi facile que ne le fait le fonctionnalisme. Pour cette raison le fonctionnalisme me semble encore une thèse plus intéressante que le physicalisme.

Hardin cherche plutôt à identifier le lien intrinsèque entre notre perception des couleurs et certaines réactions émotionnelles. Cette solution a l'avantage de pouvoir correspondre à la thèse fonctionnaliste, même si elle décrite initialement en termes physicalistes. En effet, autant les perceptions des couleurs que les réactions émotionnelles peuvent correspondre à des états fonctionnels ce qui n'était pas vraiment le cas dans la solution Shoemaker. Toutefois, cette solution a le désavantage de se fonder sur une compréhension empirique *a posteriori* de la perception des couleurs. Selon Hardin, c'est la seule manière de comprendre ce qu'est la couleur. Dans l'introduction du mémoire je me suis défendu d'entrer dans le débat du réalisme ou de l'anti-réalisme des couleurs, toutefois, ici on voit bien qu'un tel questionnement aurait des répercussions importantes sur notre débat.

Un débat plus instructif serait celui concernant le concept de nécessité. En effet, au-delà des questions de neuro-anatomie ou de réflectance de la lumière, ce que le fonctionnaliste doit montrer pour réussir à sortir de l'impasse de l'inversion du spectre, c'est que la détectabilité de l'inversion est un fait nécessaire. Il ne pourra pas montrer que c'est une nécessité *a priori* comme le désirent les anti-fonctionnalistes, mais une nécessité *a posteriori*. Cette nécessité devrait être suffisante pour sauver le fonctionnaliste. Il doit montrer que la contrainte demandée par l'anti-fonctionnaliste est trop exigeante. C'est un peu ce qu'essaie de faire Hardin quand il critique l'argument kripkéen de Levine. Hardin montre que Levine demande de la part des explications fonctionnalistes de l'esprit plus qu'il ne

demande de la part explications scientifiques concernant d'autres phénomènes. Hardin croit que le fonctionnalisme devrait être traité selon les mêmes critères que la science en général et que le scepticisme envers les explications fonctionnalistes n'est pas justifié à moins de généraliser ce scepticisme à nos autres savoirs. Ce que Hardin et Shoemaker essaient ultimement de faire c'est montrer la nécessité *a posteriori* de la détectabilité de l'inversion et que cette nécessité est tout ce qui devrait être attendu du fonctionnalisme. Toutefois, leurs tentatives ne semblent pas aussi satisfaisantes que celle de Sober.

Le téléofonctionnalisme de Sober a l'avantage de ne pas être limité à la compréhension que nous avons du cerveau. Les solutions offertes par Shoemaker et Hardin dépendent d'une compréhension particulière du cerveau humain, d'où la critique qui leur a été faite à propos de leur chauvinisme (aussi appelé « *parochialism* »). Pour que le fonctionnalisme soit une thèse intéressante, il doit se distancer d'une description qui n'aurait du sens que dans le cas de la mentalité humaine. Pour vraiment endosser la réalisation multiple, le fonctionnalisme ne doit pas être limité à notre constitution particulière. Le problème est que cette liberté se paie souvent au prix d'une absence de critères pertinents pour attribuer la mentalité. Si le fonctionnalisme ne peut pas se camper de manière stricte dans notre organisation stricte, il doit trouver un autre critère. Ce critère se trouve, selon Sober, dans le *telos*, la *raison d'être*, de l'état fonctionnel en question. Ainsi, les états fonctionnels ne doivent pas seulement être compris par ce qu'ils font et leurs relations avec d'autres états, mais aussi par ce qu'ils sont *supposés* faire. Qu'elle est la fonction intrinsèque d'un état? Sober croit que ce genre de description dissout le problème de l'inversion du spectre. Cette description est plus programmatique que décisive, mais elle a au moins le mérite de poser les fondations d'un programme de recherche intéressant. Le *telos* des états fonctionnels est, selon Sober, trouvable dans l'histoire évolutive d'un organisme. Si un état X effectue une tâche Y c'est parce qu'il est apparu à un moment particulier pour répondre à un besoin particulier. Ce genre de solution n'est pas aussi restrictif que la solution de Hardin ou autre résolution plus chauvine, mais elle reste *a posteriori* car c'est la seule manière de

comprendre l'histoire évolutive d'un organisme. Toutefois, ce genre de solution n'est pas aussi « propre » que l'espère Sober car elle pose aussi les fondations d'une pléthore de questions métaphysiques : si les états ont un *telos* y-a-t-il un plan ordonnant le tout? D'où vient ce plan? La nécessité de l'histoire évolutive doit-elle être traitée exclusivement comme une nécessité *a posteriori* ou doit-on rechercher un fondement *a priori*? Ces questions débordent bien sûr de la question de ce mémoire. Toutefois, elles doivent être envisagées par celui qui aimerait emprunter cette voie de résolution.

Outre les questions gênantes qu'elle pose, une conséquence intéressante de cette thèse est que la philosophie de l'esprit, si elle veut maintenir la thèse fonctionnaliste, doit passer par la philosophie de la biologie pour mieux comprendre ce qu'est une fonction. Le titre de l'article de Sober⁹⁴ montre très bien le problème du fonctionnalisme «classique»: il a perdu de vue (ou jamais bien établi) ce qu'est une fonction. Une précision est ici importante: je ne veux pas affirmer que le concept de fonction est compris de manière définitive en philosophie de la biologie, mais bien que, au moins, cette dernière se concentre sur la question de ce qu'est une fonction, alors que le fonctionnalisme en philosophie de l'esprit prend ce concept pour acquis. Tant que le fonctionnalisme n'examinera pas ce qu'il entend réellement par «fonction», il ne peut pas espérer répondre aux objections du genre de l'inversion du spectre.

Toutefois, le téléofonctionnalisme pourrait avoir des conséquences troublantes pour le fonctionnalisme en général. Sober n'a pas assez défini la différence entre « ce qu'une chose fait » et « ce qu'une chose doit faire », or cette distinction me semble capitale. La réponse montrera l'épiphénoménisme des qualia. Si les qualia ne correspondent pas au but intrinsèque d'une de nos structures, mais seulement à quelque chose qu'elle fait, ils doivent être considérés comme des épiphénomènes (c'est la voie que semble adopter Sober, mais le téléofonctionnalisme n'est pas lié à cette conclusion). Or un tel épiphénoménisme affaiblirait la position pro-qualia (qui est généralement anti-fonctionnaliste), mais

⁹⁴ « *Putting the Function Back into Functionalism* » voir note 81 pour référence complète.

aussi potentiellement le fonctionnalisme (et le téléofonctionnaliste) : si le qualia ne sont que des épiphénomènes, notre description verbale de ces qualia ne révélerait pas vraiment ce que fait le cerveau, ou comment fonctionne notre esprit. Or le fonctionnalisme est fondé sur le rapport offert par le comportement. Le téléofonctionnalisme est plus efficace que le fonctionnalisme car il peut expliquer pourquoi nous avons ces épiphénomènes. En utilisant l'histoire évolutive, le téléofonctionnalisme peut trouver la raison d'être de ces épiphénomènes, comme structures « vestigiales ». L'utilisation de l'histoire évolutive constitue un avantage sérieux pour le téléofonctionnalisme. Le problème est qu'en réduisant la pertinence du comportement, il est légitime de se demander si le téléofonctionnalisme peut vraiment être considéré comme une thèse fonctionnaliste. Ainsi, il n'est pas clair que le téléofonctionnalisme puisse vraiment être considéré comme une manière de sauver le fonctionnalisme.

Ainsi, le téléofonctionnalisme de Sober est une option plus intéressante que le fonctionnalisme *simpliciter* parce qu'il offre un critère de description des états qui soutient la réalisation multiple tout en restreignant son champ à l'histoire évolutive. Ce genre de description rend compte des qualia de sorte à ne pas les rendre triviaux, ou permet au moins d'expliquer cette trivialité de manière historique. Le gouffre de Levine est donc toujours là, mais le téléofonctionnalisme, sans avoir déjà la réponse, a au moins les outils pour y répondre éventuellement : on ne peut peut-être pas combler le gouffre explicatif, mais on peut possiblement expliquer pourquoi il est là, et cela est peut-être suffisant pour sauver le fonctionnalisme et le matérialisme en général.

BIBLIOGRAPHIE I

Les ouvrages suivants sont ceux qui ont directement servi à la rédaction de ce mémoire (la seconde bibliographie comporte la liste d'autres ouvrages m'ayant servi indirectement comme mise en contexte de la problématique)

- Antony, Michael V. 1994. « *Against Functional Theories of Consciousness.* » *Mind and Language* Vol.9(N.2):105-23.
- Bernier, Paul. « *Qualia Similarity and Inverted Earth.* » . Manuscrit
- . « *Self-Knowledge, Fallibility and Externalism.* » . Manuscrit
- Bickerton, Derek. 1995. *Language and Human Behavior*. Seattle: University of Washington Press.
- . 1990. *Language and Species*. Chicago: The University of Chicago Press.
- Block, Ned. 1990. « *Inverted Earth.* » *Philosophical Perspectives, Action Theory and Philosophy of Mind* v.4:53-79.
- Block, Ned, Owen Flanagan, and Güven Güzeldere, eds. 1997. *The Nature of Consciousness*. Cambridge, MA: MIT Press Bradford Book.
- Block, Ned and Jerry A. Fodor. 1972. « *What Psychological States Are Not.* » *The Philosophical Review* LXXXI(april):159-81.
- Byrne, Alex and David R. Hilbert, eds. 1997. *Readings on Color Volume 1: The Philosophy of Color*. Cambridge, MA: MIT Press (Bradford Book).
- , eds. 1997. *Readings on Color Volume 2: The Science of Color*. Cambridge, MA: MIT Press (Bradford Book).
- Chalmers, David J. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Churchland, Patricia S. 1986. *Neurophilosophy : Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press (Bradford Book).
- Churchland, Paul M. 1988. *Matter and Consciousness*. revised edition ed. Cambridge MA: MIT Press Bradford Book.
- Churchland, Paul M. and Patricia S. Churchland. 1981. « *Functionalism, Qualia, and Intentionality.* » *Philosophical Topics* XXII(1):121-45.

- Clark, Austen. 1985. « *A Physicalist Theory of Qualia.* » *The Monist* LXVIII(4):491-506.
- . 1985. « *Qualia and the Psychophysiological Explanation of Color Perception.* » *Synthese* v.65:377-405.
- . 1985. « *Spectrum Inversion and the Color Solid.* » *The Southern Journal of Philosophy* XXIII(4):431-43.
- Cole, David. 1990. « *Functionalism and Inverted Spectra.* » *Synthese* 82:207-22.
- Dedrick, Don. 1996. « *Can Colour Be Reduced to Anything.* » *Philosophy of Science* v.63:S134-S142.
- Dennett, Daniel C. 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* New York: Simon & Shuster.
- . 1988. « *Précis of the Intentional Stance.* » .
- Flanagan, Owen. 1992. *Consciousness Reconsidered.* Cambridge, MA: MIT Press (Bradford Book).
- Hammond, Constance, editor. 1996. *Cellular and Molecular Neurobiology.* San Diego: Academic Press.
- Hardin, C. L. 1988. *Color for Philosophers; Unweaving the Rainbow.* 2nd edition ed. Indianapolis: Hackett.
- . 1987. « *Qualia and Materialism: Closing the Explanatory Gap.* » *Philosophy and Phenomenological Research* XLVIII(2):281-98.
- . 1991. « *Reply to Levine.* » *Philosophical Psychology* v.4(n.1):41-50.
- . 1991. « *Reply to Teller.* » *Philosophical Psychology* v.4(n.1):61-64.
- . 1992. « *The Virtues of Illusion.* » *Philosophical Studies* v.68:371-82.
- Harding, Gregory. 1991. « *Color and the Mind-Body Problem.* » *Review of Metaphysics* v.45:289-307.
- Harman, Gilbert. 1990. « *The Intrinsic Quality of Experience.* » *Philosophical Perspectives, Action Theory and Philosophy of Mind* v.4:31-51.
- Harvey, J. 1979. « *Systematic Transposition of Colours.* » *Australasian Journal of Philosophy* v.57(n.3):211-19.
- Hilbert, David R. 1992. « *What Is Colour Vision.* » *Philosophical Studies* v.68:351-70.

- Horgan, Terence. 1984. «*Functionalism, Qualia, and the Inverted Spectrum.*» *Philosophy and Phenomenological Research* XLIV(4):453-69.
- Horowitz, Tamara and Gerald J. Massey, eds. 1991. *Thought Experiments in Science and Philosophy*. Rowman & Littlefield Publishers, Inc.
- Johnsen, Bredo C. 1993. «*The Intelligibility of Spectrum Inversion.*» *Canadian Journal of Philosophy* v.23(n.4):631-36.
- Johnsen, Bredo C. 1986. «*The Inverted Spectrum.*» *Australasian Journal of Philosophy* 64(4):471-76.
- Kernohan, Andrew. 1990. «*Lewis's Functionalism and Reductive Materialism.*» *Philosophical Psychology* v.3(n.2):235-46.
- Kim, Jaegwon. «*Reduction, Reductive Explanation, and "the Explanatory Gap".* » .
 ———. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.
- Kirk, Robert. 1981-1982. «*Goodbye to Transposed Qualia.* » *Aristotelian Society Proceedings* LXXXII:33-44.
- Kobes, Bernard W. 1991. «*Sensory Qualities and "Homunctionalism": a Review Essay of W.G. Lycan's Consciousness.* » *Philosophical Psychology* v.4(n.1):147-57.
- Kosslyn, Stephen M. and Daniel N. Osherson, eds. 1995. *Visual Cognition :An Invitation to Cognitive Science Volume 2*. second edition ed. Cambridge, MA: MIT Press (Bradford book).
- Levine, Joseph. 1988. «*Absent and Inverted Qualia Revisited.* » *Mind and Language* Vol.3(No.4):271-87.
- . 1991. «*Cool Red.* » *Philosophical Psychology* v.4(n.1):27-40.
- . 1983. «*Materialism and Qualia: The Explanatory Gap.* » *Pacific Philosophical Quarterly* 64:354-61.
- Lewis, David. 1995. «*Should a Materialist Believe in Qualia?* » *Australasian Journal of Philosophy* v.73(n.1):140-144.
- Lycan, William G. 1973. «*Inverted Spectrum.* » *Ratio* 15:315-19.
- McGinn, Colin. 1981. «*A Note on Functionalism and Function.* » *Philosophical Topics* XII(1):169-70.
- Mundale, Jennifer and William Bechtel. 1996. «*Integrating Neuroscience, Psychology, and Evolutionary Biology Through a Teleological Conception of*

- Function.* » *Minds and Machines* 6:481-505.
- Nida-Rümelin, Martine. 1998. « *On Belief About Experiences. An Epistemological Distinction Applied to the Knowledge Argument Against Physicalism.* » *Philosophy and Phenomenological Research* LVIII(1):51-73.
- . 1996. « *Pseudonormal Vision: An Actual Case of Qualia Inversion.* » *Philosophical Studies* 82:145-57.
- Rhees, Rush. 1968. « *Wittgenstein's Notes for Lectures on "Private Experience" and "Sense Data".* » *The Philosophical Review* v.77(July):271-320.
- Shear, Jonathan, ed. 1997. *Explaining Consciousness, The Hard Problem.* Cambridge, MA: MIT Press (Bradford Book).
- Sheldon Davies, Paul. 1996. « *Evolutionary Theory in Cognitive Psychology.* » *Minds and Machines* 6:445-62.
- Shoemaker, Sydney. 1980. « *Functionalism and Qualia.* » in *Readings in Philosophy of Psychology* Ed. Ned Block v.1:251-67.
- . 1982. « *The Inverted Spectrum.* » *The Journal of Philosophy* LXXIX(7):357-81.
- . 1993. « *Lovely and Suspect Ideas; Dennett Symposium.* » *Philosophy and Phenomenological Research* LIII(4):905-10.
- . 1994. « *Phenomenal Character.* » *Noûs* v.28(no.1):21-38.
- . 1990. « *Qualities and Qualia: What's in the Mind?* » *Philosophy and Phenomenological Research* L supplement:109-31.
- . 1981. « *Some Varieties of Functionalism.* » *Philosophical Topics* XII(1):93-119.
- Sorensen, A. R. 1992. *Thought Experiments.* Oxford: Oxford University Press.
- Teller, Davida Y. 1991. « *Simpler Arguments Might Work Better.* » *Philosophical Psychology* v.4(n.1):51-60.
- Tye, Michael. 1994. « *Qualia, Content, and the Inverted Spectrum.* » *Noûs* v.28(No2):159-83.

BIBLIOGRAPHIE II

Les ouvrages suivants m'ont servi indirectement comme mise en contexte de la problématique (la première bibliographie comporte la liste des ouvrages ayant directement servi à la rédaction de ce mémoire).

- Altham, J. J. 1988. « *Critical Notice of Identity, Cause, and Mind: Philosophical Essays, by Shoemaker.* » *Mind* xcii(386):285-90.
- Carruthers, Peter. 1996. *Language, Thought and Consciousness-An Essay in Philosophical Psychology*. Cambridge, England: Cambridge University Press.
- Churchland, Paul M. and Patricia S. Churchland. 1998. *On the Contrary: Critical Essays, 1987-1997*. Cambridge, MA: MIT Press Bradford Book.
- Clark, Andy. 1997. *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press Bradford Book.
- Clark, Austen. 1993. *Sensory Qualities*. Oxford: Clarendon Press.
- . 1996. « *True Theories, False Colors.* » *Philosophy of Science* v.63:S143-S150.
- Dancy, Jonathan and Ernest Sosa, eds. 1992. *A Companion to Epistemology*. Oxford: Blackwell.
- Dennett, Daniel C. 1998. *Brainchildren: Essays on Designing Minds*. Cambridge, MA: MIT Press Bradford Book.
- . 1991. « *Real Patterns.* » *The Journal of Philosophy* 88:27-51.
- Dretske, Fred. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press (Bradford Book).
- Fortin, Claudette and Robert Rousseau. 1989. *Psychologie cognitive: une approche de traitement de l'information*. Ste-Foy: Éditions de la Télé-Université.
- Goldman, Alvin I., ed. 1993. *Readings in Philosophy and Cognitive Science*. Cambridge, MA: MIT Press (Bradford Book).
- Guttenplan, Samuel, editor. 1994. *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Hall, Richard J. 1996. « *The Evolution of Color Vision Without Colors.* » *Philosophy of Science* v.63:S125-S133.

- Harrison, Bernard. 1967. « *On Describing Colours.* » *Inquiry* 10:38-52.
- Kraut, Robert. 1992. « *The Objectivity of Color and the Color Objectivity.* » *Philosophical Studies* v.68:265-87.
- Lowe, E. J. 1996. *Subjects of Experience.* Cambridge: Cambridge University Press.
- Lycan, William G. 1987. *Consciousness.* Cambridge, MA: MIT Press.
- Mack, Arien a. R. I. 1998. *Inattentional Blindness.* Cambridge, MA: MIT Press.
- McCauley, Robert N., ed. 1996. *The Churchlands and Their Critics.* Cambridge, MA: Blackwell.
- Meyer, Philippe. 1997. *L'oeil et le cerveau: biophilosophie de la perception visuelle.* Paris: Odile Jacob.
- Nelkin, Norton. 1996. *Consciousness and the Origins of Thought.* Cambridge, England: Cambridge University Press.
- Peterson, Roland and Sybe Terwee. 1994. « *Can Functionalism Provide the Proper Basis for a Core Theory of Psychoanalysis?* » *Philosophical Psychology* v.7(n.4):463-69.
- Quine, Willard V. O. 1960. *Word and Object.* Cambridge MA: MIT Press.
- Rey, Georges. 1992. « *Sensational Sentences Switched.* » *Philosophical Studies* 68:289-319.
- Sacks, Oliver. 1995. *An Anthropologist on Mars; Seven Paradoxal Tales.* London: Picador.
- Shepard, Roger N. 1992. *L'oeil qui pense: visions, illusions, perceptions.* édition française (trad. par Jacqueline Henry) ed. Paris: Seuil.
- Thagard, Paul. 1996. *Mind; Introduction to Cognitive Science.* Cambridge, MA: MIT Press.
- , editor. 1998. *Mind Readings; Introductory Selection on Cognitive Science.* Cambridge, MA: MIT Press.
- Thompson, Evan. 1992. « *Novel Colours.* » *Philosophical Studies* v.68:321-49.
- Von Eckardt, Barbara. 1993. *What Is Cognitive Science.* Cambridge MA: MIT Press Bradford Book.
- Wittgenstein, Ludwig. 1961. *Tractatus Logico-Philosophicus Suivi Investigations Philosophiques.* Paris: Gallimard Tel.

TABLE DES MATIÈRES

INTRODUCTION.....	1
1 FONCTIONNALISMES.....	7
1.1 CONTEXTE GÉNÉRAL.....	7
1.1.1 <i>L'héritier du béhaviorisme</i>	7
1.1.2 <i>Caractérisation tripartite du fonctionnalisme</i>	9
1.2 RÉALISATION MULTIPLE.....	10
1.2.1 <i>Avantages de la réalisation multiple</i>	10
1.3 FONCTIONNALISME DU SENS COMMUN (OU FONCTIONNALISME ANALYTIQUE).....	14
1.4 FONCTIONNALISME EMPIRIQUE.....	18
2 LE GOUFFRE EXPLICATIF.....	22
2.1 ARGUMENT ÉPISTÉMOLOGIQUE DE LEVINE.....	22
2.1.1 <i>Introduction</i>	22
2.1.2 <i>L'argument de Levine</i>	23
2.1.3 <i>Conclusion</i>	26
2.2 CRITIQUE DE HARDIN.....	27
2.2.1 <i>Introduction</i>	27
2.2.2 <i>Couleur et physicalisme</i>	28
2.2.3 <i>Spéculation et « opponent process theory »</i>	29
2.2.4 <i>Cellule comparatrice et inversion du spectre</i>	33
2.2.5 <i>Conclusion</i>	35
3 INVERSION INTERSUBJECTIVE ET INTRASUBJECTIVE.....	37
3.1 ARGUMENT DE SHOEMAKER EN FAVEUR DE LA POSSIBILITÉ DE L'INVERSION.....	37
3.1.1 <i>Introduction</i>	37
3.1.2 <i>L'argument métaphysique de Shoemaker</i>	37
3.1.3 <i>Inversion intrapersonnelle ou inversion interpersonnelle</i>	40
3.1.4 <i>1^{ère} objection à la possibilité d'inversion du spectre: différence qualitative</i>	43
3.1.5 <i>2^{ème} objection à la possibilité d'inversion du spectre: incompréhension plutôt qu'inversion</i>	45
3.1.6 <i>3^{ème} objection: possibilité concrète d'une inversion</i>	47
3.2 LE NOUVEAU FONCTIONNALISME DE SHOEMAKER.....	48
3.2.1 <i>Parallèles vérificationnistes</i>	48
3.2.2 <i>Réalisation physique</i>	49
3.2.3 <i>Conclusion</i>	51
3.3 CRITIQUE DE LEVINE.....	54
3.3.1 <i>Introduction</i>	54
3.3.2 <i>Confusion entre différents types d'inversion chez Shoemaker</i>	55
3.3.3 <i>Le scepticisme dans les cas de qualia absents et d'inversion du spectre</i>	58
3.3.4 <i>Fausse concessions envers la position pro-qualia</i>	60
3.4 CONCLUSION.....	63
4 COOL RED.....	65
4.1 ARGUMENT DE LEVINE.....	65
4.1.1 <i>Introduction</i>	65
4.1.2 <i>L'expérience de pensée telle qu'envisagée par Hardin</i>	65
4.1.3 <i>Le fonctionnalisme chez Hardin</i>	69
4.1.4 <i>D'autres cas d'inversion du spectre et le compromis de Levine</i>	70
4.1.5 <i>Problèmes du matérialisme en général</i>	77
4.2 CRITIQUE DE HARDIN.....	79

4.2.1	<i>Épiphénoménisme des qualia</i>	79
4.2.2	<i>L'implausibilité de l'inversion de Levine</i>	80
4.2.3	<i>Résidu qualitatif</i>	82
4.2.4	<i>Conclusion</i>	83
5	TÉLÉOFONCTIONNALISME	86
4.3	ARGUMENT DE SOBER.....	86
5.1.1	<i>Introduction</i>	86
5.1.2	<i>Fonction biologique plutôt que mathématique</i>	87
5.1.3	<i>Le téléofonctionnalisme et les descriptions a priori</i>	90
5.1.4	<i>Épiphénoménisme des qualia</i>	93
5.1.5	<i>Avantages du téléofonctionnalisme</i>	94
6	CONCLUSION	96
	BIBLIOGRAPHIE I	101
	BIBLIOGRAPHIE II	105