

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DÉFENSE D'UNE DÉFINITION
SOCIALEMENT CONSTRUITE ET OBJECTIVE
DU CONCEPT DE TROUBLE MENTAL

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN PHILOSOPHIE

PAR ANNE-MARIE GAGNÉ-JULIEN

JUIN 2021

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je voudrais tout d'abord remercier grandement mes codirecteurs, Luc Faucher et Maël Lemoine. Je n'aurais pu rêver d'une meilleure co-supervision pour cette thèse. C'est grâce à vos avis éclairés, à vos encouragements et à votre soutien que j'ai pu mener à bien ce projet.

Merci Luc de m'avoir poussée à reformuler, à préciser et à dépasser mes idées grâce à tes commentaires toujours pointus et rigoureux. Merci de m'avoir dit patiemment que ce n'était pas suffisant quand ça ne l'était pas. Merci d'avoir été exigeant : j'ai gagné une rigueur que je n'avais pas lorsque je me suis lancée dans cette thèse, et je te le dois.

Merci Maël pour ton enthousiasme et ton ouverture face à mes intuitions parfois plus qu'embryonnaires. Merci d'avoir pris le temps de discuter avec moi pour me permettre de transformer ces intuitions en arguments. Merci pour tes suggestions qui visaient toujours à améliorer mon travail en respectant le prisme de celui-ci. Merci pour ta grande disponibilité et ton écoute. Merci de m'avoir dit que cette thèse valait la peine d'être défendue.

Merci Christophe d'avoir mis en page cette thèse et de lui avoir trouvé un titre plus digeste. Merci Pierre-Olivier pour tous tes conseils et relectures, ainsi que pour les discussions que nous avons eues, surtout celle où tu m'as parlé de la Fondation Brocher, en Suisse. Merci au Laboratoire étudiant interuniversitaire en philosophie des sciences (Montréal) pour tous les commentaires et suggestions de ses membres sur mon travail.

Merci Jean-Baptiste pour ton soutien inébranlable, ta compréhension dans mes moments de grande excitation ou de grand découragement, tes blagues qui dédramatisent et tes nombreuses relectures. Merci à ma famille d'avoir toujours cru que je pouvais finir cette thèse alors que je n'y croyais pas complètement moi-même, et de n'avoir jamais remis en question la valeur du projet. Merci à mes ami.e.s philosophes, Ely, Antoine, Geneviève, François, David, Charlotte, Simon, Gabriel, Étienne et Ghyslain, pour les discussions stimulantes et les mains tendues dans les moments plus difficiles. Merci à mes ami.e.s non philosophes de m'avoir rappelé qu'un monde existe en dehors de la thèse et qu'il vaut la peine d'aller le visiter de temps en temps.

Merci au Centre interuniversitaire de recherche sur la science et la technologie (UQAM) et à la Fondation Brocher pour les environnements de travail idéaux à la rédaction d'une thèse dont j'ai pu bénéficier. Un merci particulier à Martine, Daniel, Ana et Anyck. Merci au Conseil de recherche en sciences humaines du Canada (CRSH), au Centre interuniversitaire de recherche sur la science et la technologie (UQAM), au Groupe de recherche interdisciplinaire sur la normativité (GRIN) et à la Faculté des sciences humaines de l'UQAM pour leur soutien financier.

J'ai eu la chance d'être entourée de personnes merveilleuses tout au long de ce parcours, et il est clair que je n'aurais pas pu me rendre jusqu'au bout sans vous.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
LISTE DES FIGURES.....	x
LISTE DES ABRÉVIATIONS.....	xi
RÉSUMÉ	xii
ABSTRACT	xiii
INTRODUCTION GÉNÉRALE	1
Introduction : qu'est-ce que le trouble mental?.....	1
La crise de la légitimité de la psychiatrie et l'urgence de définir le concept de trouble mental.....	6
La question de la définition du trouble mental aujourd'hui	13
Les définitions formelles du désordre mental à partir du DSM-III.....	16
La question de la définition du trouble mental sous l'angle philosophique : deux grandes approches	25
Présentation générale de la thèse.....	34
Références de l'introduction	37
CHAPITRE 1	
LA THÉORIE BIOSTATISTIQUE DE CHRISTOPHER BOORSE	46
Introduction	46
1.1 Présentation générale de la théorie bio-statistique de Boorse	48
1.1.1 Le concept de fonction normale chez Boorse	49
1.1.2 Normalité statistique et classe de référence chez Boorse	51

1.1.3	Le naturalisme méthodologique de Boorse.....	53
1.2	L'application de la BST à la psychiatrie	58
1.2.1	Le contexte de la publication de l'article	61
1.2.2	Que devrait être une théorie de la santé mentale?.....	65
1.2.3	La thèse boorséenne de l'absence d'un concept théorique de santé mentale dans la psychiatrie des années 1970	68
1.2.4	En dehors du naturalisme méthodologique : l'attitude prescriptive de Boorse envers la psychiatrie.....	74
1.2.5	La psychanalyse comme meilleure option disponible pour élaborer une théorie du fonctionnement mental normal	77
1.2.6	La BST appliquée au DSM	80
	Conclusion	86
	Références du chapitre 1	88

CHAPITRE 2

	OBJECTIVITÉ ET INDÉPENDANCE DES VALEURS DANS LA BST : LES CRITIQUES.....	93
	Introduction.....	93
2.1	Les notions d'objectivité et de neutralité dans la BST	94
2.1.1	Premier argument : la nature empirique des concepts de santé et de pathologie.....	96
2.1.2	Deuxième argument : la neutralité de la science.....	98
2.2	Boorse contre Boorse : les valeurs admises dans la BST.....	100
2.2.1	Indétermination des buts de l'organisme et pluralisme des attributions fonctionnelles	101
2.2.2	Une défense de la BST comme une conception <i>descriptive</i> du concept théorique de pathologie émanant de la méthode de la reconstruction rationnelle 104	
2.2.3	Les choix des buts de l'organisme	106
2.2.4	Le seuil pathologique et le problème de la démarcation.....	112
2.3	Les classes de référence et l'indépendance des valeurs de la BST	117
2.3.1	Le problème de la classe de référence.....	118
2.3.2	Les classes de référence et les pratiques de classification médicale...	121

2.3.3	Les classes de référence de la BST et les critiques du dualisme homme-femme	125
	Conclusion : la présence de valeurs au sein de la BST	131
	Références du chapitre 2	133
CHAPITRE 3		
LES APPROCHES ALTERNATIVES		
	Introduction	136
3.1	Se tourner vers le constructivisme?	137
3.1.1	Présentation de l'approche constructiviste	137
3.1.2	Présentation des problèmes associés au constructivisme	146
3.2	Se tourner vers la théorie de la dysfonction préjudiciable?	151
3.2.1	La notion de dysfonction et la théorie de la sélection naturelle	153
3.2.2	L'indépendance des valeurs de la notion de dysfonction	155
3.2.3	La DP, l'essentialisme et la boîte noire	161
3.2.4	La DP et la notion de préjudice	164
3.3	Présentation des critiques de la DP	167
3.3.1	Critiques de l'approche évolutionniste	168
3.3.2	Le problème de la démarcation	175
3.3.3	Critique des implications pratiques de la DP	177
3.3.4	Critique de la notion de préjudice	179
	Conclusion	183
	Références du chapitre 3	185
CHAPITRE 4		
L'IDÉAL DE LA SCIENCE INDÉPENDANTE DES VALEURS ET LA DÉFINITION DU TROUBLE MENTAL		
	Introduction	191
4.1	L'approche objectiviste des fonctions naturelles et l'idéal de la science indépendante des valeurs	192
4.2	La conception traditionnelle de la science : l'idéal de la science indépendante des valeurs et l'objectivité comme neutralité	195

4.3	La philosophie féministe des sciences et le concept d'objectivité	198
4.4	Les différentes critiques de l'idéal de la science indépendante des valeurs 203	
4.4.1	La critique sémantique	204
4.4.2	Heather Douglas et la critique méthodologique.....	207
4.5	La critique méthodologique et l'empirisme contextualiste critique d'Helen Longino	212
4.6	Pour un nouvel idéal : l'objectivité sociale (OS).....	221
4.7	Implications des critiques de l'idéal de la science indépendante des valeurs et de l'OS pour la définition du trouble mental.....	225
4.7.1	Racisme et trouble mental.....	226
4.7.2	Sexisme et trouble mental	228
	Conclusion	231
	Références du chapitre 4.....	232

CHAPITRE 5

LIMITES DE L'OBJECTIVITÉ SOCIALE ET PROPOSITIONS

D'AMENDEMENTS.....	239
Introduction.....	239
5.1 L'OS est-elle antiféministe?.....	240
5.2 Quelle diversité pour la communauté épistémique? La théorie de la connaissance située contre l'empirisme féministe?	244
5.2.1 Présentation de la théorie féministe du <i>standpoint</i>	246
5.2.2 Convergence de la théorie du <i>standpoint</i> et de l'empirisme féministe.....	253
5.3 L'objectivité sociale et l'intégration de communautés non scientifiques ..	257
5.4 Relativisme et objectivité sociale	262
5.4.1 Précisions sur la menace du relativisme.....	265
5.4.2 Objectivité sociale et empirisme : l'adéquation empirique comme standard de justification nécessaire	267
5.4.3. Contextualisme, objectivité sociale et relativisme épistémique.....	271
5.5 Analyse de la signification de l'objectivité scientifique : objectivité métaphysique, objectivité empirique et objectivité méthodologique.....	278

Conclusion	284
Références du chapitre 5	286

CHAPITRE 6

OBJECTIVITÉ EMPIRIQUE ET THÉORIE DE LA FONCTION COMME RÔLE CAUSAL.....	292
Introduction.....	292
6.1. Les théories de la fonction comme rôle causal et le concept de trouble mental.....	294
6.1.1 La théorie de la fonction RC	298
6.1.2 Critiques de l'approche RC.....	300
6.2 La solution de Craver : attributions fonctionnelles, mécanismes causaux et perspectivalisme.....	305
6.3 La solution de Roe et Murphy : attributions fonctionnelles et idéalizations	315
6.4 Les attributions fonctionnelles et les valeurs en science	322
6.5 Objectivité empirique et mécanismes causaux.....	324
Conclusion : fonctions comme rôle causal mécanique et perspectivalisme	327
Références du chapitre 6	329

CHAPITRE 7

OBJECTIVITÉ SOCIALE ET DÉFINITION DU TROUBLE MENTAL : PROPOSITION DE CRÉATION D'UN PANEL AU SEIN DU DSM	333
7.1 Structures de prise de décision dans le DSM-5.....	334
7.2 Le processus de définition du trouble mental dans le DSM-5	338
7.3 La définition du trouble mental et les standards de l'objectivité sociale ...	345
7.3.1 Critique des processus de prises de décisions quant à la définition du trouble mental grâce au cadre de l'objectivité sociale	346
7.3.2 L'apport de T.K. Browne : un panel d'éthique responsable des valeurs qui influencent la catégorisation psychiatrique pour le DSM-5.....	350
7.3.3 Analyse de la proposition de Browne	354
7.4.1 Le panel et l'industrie pharmaceutique	357

Conclusion 363
Références du chapitre 7 366

CONCLUSION GÉNÉRALE

UN NORMATIVISME OBJECTIF POUR DÉFINIR LE TROUBLE MENTAL .. 375

BIBLIOGRAPHIE 382

LISTE DES FIGURES

Figure 1. Courbe de la distribution de l'efficacité des fonctions réalisées 114

Figure 2. Récepteur NMDA 307

LISTE DES ABRÉVIATIONS

APA	Association de psychiatrie américaine
DSM	Manuel diagnostique et statistique des troubles mentaux
ECC	Empirisme contextualiste critique
MRC	Approche mécanique du rôle causal
OFN	Approche objectiviste des fonctions naturelles
OS	Objectivité sociale
OS*	Objectivité sociale amendée
RC	Théorie de la fonction comme rôle causal

RÉSUMÉ

Dans cette thèse, je m'intéresse à l'un des enjeux centraux de la controverse sur la scientificité de la psychiatrie : la place et le rôle des valeurs sociales et culturelles dans la définition du concept de trouble mental. Plus particulièrement, je m'intéresse à la définition du trouble mental publiée dans le *Manuel diagnostique et statistique des troubles mentaux* (DSM). Contre les détracteurs de l'objectivité du concept, je défends l'idée que le trouble mental peut être à la fois socialement construit *et* objectif. J'utilise un cadre conceptuel emprunté à l'épistémologie sociale et à la philosophie féministe des sciences, soit l'empirisme contextuel critique. Développé par Helen Longino (1990, 2002), ce cadre pose que les valeurs peuvent légitimement entrer dans le processus de production de connaissances scientifiques objectives s'il existe des structures qui permettent le dialogue critique entre les différents membres d'une communauté scientifique. Cette façon de concevoir l'objectivité scientifique se nomme *objectivité sociale* et attire l'attention sur les processus sociaux qui permettent la production de connaissances scientifiques. Suivant cette approche, je propose que le concept de trouble mental puisse être objectif si sa définition permettait l'intégration de la critique des différents groupes concernés par cette définition (groupes de patients, psychiatres, etc.).

ABSTRACT

In this dissertation, I examine an issue at the heart of the controversy regarding the scientific status of psychiatry, namely the place and roles of social and cultural values in the definition of the mental disorder concept. In particular, I investigate the definition of mental disorder published in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM). Against those who contest this concept's objectivity, I contend that mental disorder can be both *socially constructed* as well as *objective*. To this end, I borrow a conceptual framework from social epistemology and feminist philosophy: critical contextual empiricism. Developed by Helen Longino (1990, 2002), this framework posits that values can legitimately enter the process of producing objective scientific knowledge if there are structures which allow critical dialogue among the various members of a scientific community. This conception of scientific objectivity is called *social objectivity*, as it highlights the social processes that enable the production of scientific knowledge. Taking this approach, I propose that the concept of mental disorder could be objective if the way it is defined incorporates critiques by the different groups affected by this definition (patients, psychiatrists, etc.).

INTRODUCTION GÉNÉRALE

Dans cette thèse, je m'intéresserai à la place et aux rôles des valeurs sociales et culturelles dans la définition du concept de trouble mental. Contre les critiques du concept, je défendrai que le trouble mental peut être à la fois socialement construit et objectif. J'utiliserai un cadre conceptuel emprunté à l'épistémologie sociale et à la philosophie féministe des sciences, soit « l'empirisme contextuel critique », pour penser l'objectivité du concept de trouble mental. Je défends que le concept de trouble mental peut être objectif si la manière dont sa définition est élaborée permet l'intégration de la critique des différents groupes touchés par cette définition (patients, chercheurs en sciences humaines, psychiatres et autres professionnels de la santé mentale, décideurs publics, etc.).

Introduction : qu'est-ce que le trouble mental?

L'histoire de la psychiatrie est émaillée d'une longue liste de controverses portant sur ses catégories diagnostiques. En 1851, Samuel Cartwright inventait le diagnostic de drapétomanie pour désigner les esclaves qui tentent de fuir leur maître (Gamwell et Tomes, 1995). Jusqu'en 1973, l'homosexualité était considérée comme un trouble

mental par le DSM¹ (p. ex. : Kirk et Kutchins, 1992). Ces quelques exemples historiques, parmi tant d'autres, de pathologisation de populations marginalisées, font ombrage aux prétentions scientifiques et objectives de la psychiatrie contemporaine. Récemment encore, l'inclusion du trouble oppositionnel avec provocation (TOP) au sein du DSM a été le sujet de vifs débats (Potter, 2012, 2014). Considérant que l'un des symptômes centraux du TOP est l'opposition et la défiance vis-à-vis l'autorité, ne serait-il pas possible que cette condition ne soit pas un réel trouble mental, mais plutôt un comportement dévalorisé ou dérangeant? L'inclusion de plusieurs autres conditions au sein du DSM ne fait pas consensus. C'est le cas du trouble bipolaire, des troubles d'apprentissage, du trouble d'anxiété sociale, du trouble de déficit de l'attention avec hyperactivité (TDAH), des dysfonctions sexuelles et des paraphilies, du trouble d'identité de genre et du syndrome dysphorique prémenstruel (Lane, 2007; Wakefield et First, 2003). Même des diagnostics paradigmatiques comme la schizophrénie (Heinrichs, 2001; Rosenhan, 1973) ou la dépression majeure (Horwitz et Wakefield, 2007) ne font pas consensus. Tout positionnement vis-à-vis ces différents diagnostics relève d'une même question : qu'est-ce que le trouble mental? Cette question entraîne un vaste lot d'interrogations connexes (Radden, 2019; Schuur 2020).

La dépression majeure et la schizophrénie sont-elles de même nature que le cancer ou le diabète? L'esprit peut-il être malade au même titre que le corps? La spécificité des troubles *mentaux* par rapport aux pathologies traitées par la médecine somatique est aussi un point de litige, d'autant plus que le DSM a l'ambition de traiter des troubles dont les symptômes sont associés à des dysfonctions neurophysiologiques. Or ces *patterns* neurophysiologiques semblent extrêmement difficiles à identifier (p. ex. : Sykes, 2010). Par ailleurs, certaines approches philosophiques conçoivent le trouble mental dans une perspective phénoménologique, en termes de représentation et de

¹ Le sigle DSM réfère au *Diagnostic and statistical Manual of Mental Disorders* produit par l'American Psychiatric Association (ci-après APA).

capacité d'action (p. ex. : Parnas et Sass, 2008; Tekin 2019; Thornton, 2007), alors que d'autres le comprennent dans une perspective réductionniste, comme une dérégulation neurophysiologique (p. ex. : Murphy, 2006).

Au-delà de la question de la distinction entre le mental et le physique, celle de la distinction entre ce qui est un trouble et ce qui ne l'est pas est également cruciale. Quel est le rapport entre les troubles mentaux et ce que l'on pourrait appeler des « défauts » (p. ex. : la timidité) ou des « vices » (p. ex. : la psychopathie), voire des problèmes rencontrés par chacun au cours de son existence (p. ex. : l'anxiété, le deuil)? Le trouble mental est-il une forme d'irrationalité, un désordre de l'agentivité, un handicap lié à une mauvaise adéquation avec l'environnement social, une dysfonction neurophysiologique par rapport au fonctionnement naturel de l'esprit humain ou un état de souffrance et de détresse nécessitant une aide médicale? Et si l'on accepte l'une ou l'autre de ces conceptualisations du trouble mental, encore faut-il la préciser, car des notions comme « détresse » ou « dysfonction » peuvent être elles-mêmes polysémiques. Qui plus est, si l'on accepte que le concept de trouble mental soit défini par des critères comme l'inadéquation avec l'environnement ou le vécu subjectif des individus, est-ce que cela implique que la distinction entre le normal et le pathologique est à trouver du côté du social, et non du biologique? Et si, en répondant positivement à cette question, il est admis que les valeurs font partie du processus de classification et de diagnostic en psychiatrie, cette discipline peut-elle prétendre au titre de science?

Le trouble mental est-il le simple opposé de la santé mentale? Certaines approches définissent la santé comme l'absence de maladie. Dans cette optique, si vous êtes en santé, vous n'êtes pas malade et si vous êtes malade, vous n'êtes pas en santé (Hofmann 2005, p. 357). Or si l'on pense, comme l'Organisation mondiale de la santé (OMS), que la santé est « un état de complet bien-être physique, mental et social, et ne consiste pas seulement en une absence de maladie ou d'infirmité » (1946, non modifiée depuis),

la relation entre les deux concepts apparaît plus complexe. En plus de l'absence de maladie, la définition de l'OMS demande un idéal beaucoup plus exigeant, celui du bien-être. Dès lors, si vous n'êtes pas malade, vous n'êtes pas nécessairement en santé. De plus, y a-t-il des distinctions à faire entre la « pathologie », qui serait un état de désordre neurophysiologique, et la « maladie », qui serait un état de détresse ou d'incapacité vécu et subjectif? Peut-il y avoir plus d'une définition du trouble mental, en fonction des cultures, des sociétés et des contextes scientifiques dans lesquels le terme est utilisé? Le concept de trouble mental utilisé dans le cadre légal pour juger de la responsabilité d'un accusé est-il le même que celui du psychiatre qui accompagne son patient en clinique?

L'ensemble de ces questions a conduit les philosophes de la psychiatrie à tenter de définir formellement le trouble mental. Néanmoins, déterminer la bonne définition du trouble mental soulève la controverse depuis maintenant plusieurs décennies, selon que les auteurs adoptent des perspectives biomédicale, psychocognitive, comportementale, psychanalytique ou sociale ou que les auteurs s'intéressent plus ou moins à chacune des sous-questions abordées plus haut (Ghaemi, 2004; Tyrer et Steinberg, 2006). En parallèle à ces questionnements philosophiques, depuis la parution du DSM-III, l'APA s'est par ailleurs dotée d'une définition formelle du trouble mental, qui a elle-même été discutée et critiquée.

Mon objectif dans cette thèse est de prendre part au débat sur la définition du trouble mental. Or comme le suggère l'ensemble des enjeux énumérés ci-haut, la question « qu'est-ce que le trouble mental? » est vaste et complexe. Il convient donc de restreindre l'analyse à certains aspects de la question. Dans cette thèse, je m'intéresserai plus particulièrement à la place et aux rôles des valeurs sociales et culturelles dans la définition du trouble mental offerte par le DSM. Comme nous le verrons, l'influence des valeurs sur la psychiatrie et sa définition du trouble mental est

un thème central à la fois pour la psychiatrie et pour la philosophie de la psychiatrie. C'est également un thème dont les présuppositions méritent d'être remises en question. Plus spécifiquement, la distinction entre faits et valeurs et la notion d'objectivité scientifique seront au cœur de l'investigation. Qui plus est, même si la question de la définition du trouble mental pourrait être posée pour elle-même, je m'intéresserai particulièrement au lien entre le débat philosophique et la conception du trouble mental qui se dégage du DSM, et ce, pour deux raisons. D'abord, le DSM est un outil extrêmement influent dans le domaine des soins et de la recherche en santé mentale en Amérique du Nord. Les cliniciens doivent utiliser le manuel pour poser le diagnostic, et les patients ont quant à eux généralement besoin d'un diagnostic du DSM pour recevoir des thérapies particulières et un remboursement des compagnies d'assurances. La vaste majorité de la recherche en santé mentale doit faire correspondre son vocabulaire à celui du DSM pour pouvoir être financée et publiée. En raison de cette grande influence, la conception du trouble mental du DSM me semble un objet d'étude important et mérite que je m'y intéresse spécifiquement. Ensuite, historiquement, le débat philosophique, du moins dans sa version anglo-saxonne et analytique, a émergé suite aux controverses affrontées par la psychiatrie américaine et son outil diagnostique. En ce sens, il me semble opportun de conserver cette attention particulière portée au DSM pour m'assurer que le débat philosophique ne perde pas de vue les enjeux concrets et pratiques d'une telle question. Qui plus est, étant donné que je m'intéresse particulièrement à la définition du trouble mental du DSM, je ne vais pas m'attarder à une analyse de la santé mentale et de la relation entre santé mentale et trouble mental. Ce choix s'explique par le fait que ces dimensions du débat ne semblent pas prioritaires pour le DSM.

Cette thèse se présente donc comme une défense de l'idée selon laquelle le trouble mental peut être à la fois socialement construit et objectif, en adoptant le cadre théorique de *l'objectivité sociale*, développé en épistémologie sociale des sciences.

Dans ce qui suit, pour mieux faire émerger comment le thème de la distinction entre faits et valeurs s'impose à une analyse de la définition du trouble mental d'un point de vue philosophique, je procéderai en quatre étapes : 1) reconstruction du contexte historique et social de l'émergence de la définition formelle du trouble mental dans le DSM, 2) analyse de l'évolution de cette définition dans les DSM ultérieurs, 3) mise en lumière des enjeux de définition du trouble qui animent le débat philosophique traditionnel et 4) exploration d'une piste de solution pour penser la question de la définition du trouble mental, l'épistémologie sociale. Pour clore cette introduction, je présenterai brièvement l'organisation générale de la thèse et le contenu de chacun de ses chapitres.

La crise de la légitimité de la psychiatrie et l'urgence de définir le concept de trouble mental

Pour bien situer mon projet de thèse, il convient de mettre en évidence le contexte au sein duquel les psychiatres ont tenté de définir formellement le concept de trouble mental et pourquoi cette tâche leur a paru essentielle. C'est parce que la psychiatrie américaine traversait une crise de légitimité sans précédent que l'APA a été forcée de définir le trouble mental. Cette période historique agitée a reçu beaucoup d'attention de la part des historien.nes et philosophes de la psychiatrie (p. ex. : Decker, 2013; Demazeux, 2013, 2021; Grob, 2011; Porter, 2002; Shorter, 1997, 2013). Deux facteurs principaux ont mené à l'élaboration de cette définition : les critiques du mouvement antipsychiatrique, selon lequel le trouble mental est un mythe, et la crise entourant la classification de l'homosexualité comme un trouble mental. Retraçons les grandes

lignes de ce moment charnière de la psychiatrie américaine, avec comme fil conducteur les différentes conceptions du trouble mental qui s'y affrontaient.

Les deux premières éditions du DSM, les DSM-I et DSM-II, publiées respectivement en 1952 et en 1968, étaient des manuels relativement courts s'inscrivant dans une approche psychosociale et psychanalytique. Ces deux premiers DSM ne contenaient pas de définition formelle du trouble mental. Malgré l'absence de définition, il est tout de même possible de faire émerger la signification du concept. Selon l'approche psychanalytique, les conditions psychiatriques sont généralement conceptualisées comme l'expression de tensions sous-jacentes aux symptômes, causées par des événements problématiques qu'affronte l'individu. Selon l'historienne E. Wilson, le trouble mental est caractérisé au sein du cadre psychanalytique par quatre principaux éléments :

1) [...] the boundary between the mentally well and the mentally ill is fluid because normal persons can become ill if exposed to severe-enough trauma, 2) [...] mental illness is conceived along a continuum of severity-from neurosis to borderline conditions to psychosis, 3) [...] an untoward mixture of noxious environment and psychic conflict causes mental illness, and 4) [...] the mechanisms by which mental illness emerges in the individual are psychologically mediated (known as the principle of psychogenesis) (Wilson, 1993, p. 400).

Ainsi, le modèle psychanalytique présente la frontière entre les troubles mentaux et les états normaux comme floue, ces troubles étant causés par une mauvaise adaptation entre l'individu et son environnement, résultant d'une trajectoire individuelle unique. En raison de l'unicité de ces conditions et de l'absence de différence qualitative entre ce qui est un désordre et ce qui ne l'est pas, le diagnostic ne joue donc pas un rôle

central au sein de la psychanalyse. C'est pourquoi une définition formelle du trouble mental ne semblait pas cruciale à la pratique psychiatrique².

C'est au tournant des années 1950 que la psychiatrie d'orientation psychanalytique s'est vue devenir l'objet de virulentes critiques qui ont ultimement mené à sa transformation vers le modèle médical, centré sur le diagnostic. Ces différentes attaques étaient hétérogènes et provenaient à la fois d'acteurs internes et externes à cette discipline. Plusieurs facteurs ont mené la psychiatrie à se métamorphoser radicalement à cette période (Mayes et Horwitz, 2005). Les critiques visaient principalement l'approche thérapeutique promue par la psychanalyse, la psychothérapie, qui ne semblait pas donner les résultats escomptés en termes de rémission des symptômes, problème de taille pour les compagnies d'assurances (Cooper, 2005; Mayes et Horwitz, 2005; Shorter, 2013; Tsou 2011). Les mauvaises conditions de soins et d'hébergement dans les asiles psychiatriques ont également été révélées à un public désormais choqué (Goffman, 1961; Grob, 1994; Jones, 1993; on pense aussi, entre autres, au film *Vol au-dessus d'un nid de coucou*, réalisé par Forman en 1975). Le caractère scientifique des fondements et de la méthode psychiatrique ainsi que la validité des construits diagnostiques étaient également remis en question (Decker, 2007; Jablensky, 2007). Ces critiques, bien que ciblant différents aspects de la pratique et de la théorie psychiatrique, ont fait émerger le besoin pour cette discipline et pour son outil diagnostique principal, le DSM, de défendre son objectivité et son caractère descriptif, sans quoi elle ne pourrait affirmer son statut de sous-discipline de la médecine. C'est dans ce contexte que la troisième édition du DSM a été élaborée.

² Mentionnons toutefois que l'histoire de la psychanalyse est beaucoup plus riche que ce que ce court paragraphe la fait paraître, ce qui ne rend pas justice aux différentes approches psychanalytiques sur le trouble mental. Pour une reconstruction des courants diversifiés de la psychanalyse, voir Porter 2002, chapitre 8. Je reviens sur la conceptualisation du trouble mental dans les théories psychanalytiques dans le chapitre 1.

Face à cet ensemble diversifié de critiques, les objectifs du DSM-III en termes de scientificité peuvent être découpés en trois grands points, suivant le philosophe D. Bolton :

- 1) Diagnosis should be reliable, with adequate agreement between clinicians, or between raters in research studies, as to what conditions of interest are and when they are present or absent.
- 2) Classification should be valid, both classification of symptoms into syndromes, and syndromes into higher-level categories along with other syndromes. [...]
- 3) The manuals should capture all and only mental disorders, excluding current false positives (non-disorders currently in the manuals), and bringing in current false negatives (disorders not yet included). This can be expressed as the aim of making the manuals more valid as manuals of mental disorders. (Bolton, 2008, p. 29)

L'objectif 1 concerne davantage les enjeux de fiabilité des catégories et de facilitation de la communication. L'objectif 2 renvoie à la validité des catégories diagnostiques spécifiques (référant à des questionnements de la forme « le critère X est-il nécessaire à l'identification du désordre Y »?) et des classes de syndromes (référant à des questionnements de la forme « le désordre X devrait-il être inclus dans la classe de syndromes Y ou Z »?). Ces objectifs sont donc distincts de l'objectif 3, qui concerne spécifiquement le trouble mental en tant que catégorie générale recoupant tous les syndromes et classes de syndromes particuliers. Il ne faut donc pas confondre les problèmes entourant les définitions des troubles spécifiques (2) et la définition de la catégorie générale de trouble, qui inclue les troubles spécifiques (3) (voir Sadegh-Zadeh, 2008). Au regard du point 3, deux critiques plus spécifiques ont mené à la transformation de la psychiatrie psychanalytique vers une psychiatrie biomédicale sur le plan de la conceptualisation du trouble mental. Il s'agit de la critique des antipsychiatres quant à la validité du concept de trouble mental (p. ex. : Szasz, 1961) et la crise entourant le « trouble » de l'homosexualité dans les DSM-I et II (Bayer, 1981, 1987; Kirk et Kutchins 1992; Kutchins et Kirk 1997). C'est face à ces deux

critiques que s'est élaborée la première définition formelle du désordre mental dans le DSM-III.

Bien que le mouvement antipsychiatrique soit hétérogène et porté par des auteurs de diverses allégeances théoriques (p. ex. : Laing et Esterton, 1964; Rosenhan, 1963; Szasz, 1961; voir Garson, 2006 pour une discussion), l'une des critiques vers laquelle ce mouvement converge au regard du trouble mental est la suivante : la psychiatrie serait une institution dont la fonction centrale est de réguler les comportements socialement déviants. L'attaque est cependant incarnée par l'un des plus grands représentants du mouvement antipsychiatrique : Thomas Szasz (1960, 1961). Pour Szasz, le trouble mental n'est qu'un mythe, en rien lié au cadre médical dans lequel la psychiatrie souhaitait s'intégrer. Cette attaque se fonde d'abord sur une première observation : la difficulté pour la psychiatrie de valider son concept de trouble mental sur une base physiologique. Elle est renforcée par le postulat psychosocial selon lequel un trouble mental résulte d'une incompatibilité entre l'individu et son environnement. Comme l'a bien mis en lumière S. Varga, cette critique repose sur deux prémisses sous-jacentes : 1) les troubles possèdent des bases physiologiques en ce qu'ils sont des dysfonctions du corps humain et 2) les troubles peuvent être objectivement identifiés en tant que déviations par rapport à des normes physiologiques fonctionnelles (Varga, 2015, p.108; voir Kingma, 2013 pour une reconstruction de l'argument spécifiquement Szaszien). Étant donné l'absence de frontière physiologique qui pourrait légitimer l'attribution d'un désordre mental à un individu, cette frontière ne pourrait être que sociale. Cette approche nie donc l'existence des troubles mentaux comme catégories médicales parce qu'elles ne posséderaient pas de référents physiologiques. La psychiatrie serait donc un processus d'exclusion sociale, où attribuer à un individu un trouble mental signifierait que son état psychologique ou comportemental n'est pas conforme aux valeurs de la société dans laquelle il se trouve. La critique relativiste des antipsychiatres peut être résumée comme suit :

- 1) Les normes, qui sont non objectives, dépendent du contexte social;
- 2) L'attribution de trouble dépend de normes;
- 3) Donc, parce que l'attribution de troubles dépend du contexte social, cette attribution est non objective.

Cette relativité expliquerait donc pourquoi des états sains, mais jugés hors norme par la société, seraient désignés comme pathologiques (Garson 2006; Kirk et Kutchins, 1992; Mayes et Horwitz, 2005; Varga 2015; Wilson, 1993). La psychiatrie ne pouvant démontrer que les troubles mentaux sont des dysfonctions de nature physiologique, elle ne peut donc pas prouver l'existence des troubles mentaux, ce pour quoi elle a été qualifiée de processus d'exclusion sociale.

La crise entourant le diagnostic de l'homosexualité est généralement perçue comme le second facteur majeur ayant poussé l'APA à définir le trouble mental (p. ex. : Bayer, 1981, 1987; Drescher et Merlino, 2007; Kirk et Kutchins, 1992; Kutchins et Kirk, 1997). C'est aussi une manifestation du flou entourant le concept de trouble mental du DSM de l'époque. Dans les DSM I et II, l'homosexualité était considérée comme une forme pathologique de développement sexuel. À partir des années 1970, de nombreux groupes de pression militant pour les droits des personnes homosexuelles ont revendiqué la normalité de cette orientation. Par exemple, la réunion annuelle de l'APA de 1970 et plusieurs autres symposiums ont été perturbés par des manifestations en faveur des droits des homosexuels (p. ex. : Cooper, 2005). Des psychiatres homosexuels ont également revendiqué la normalité de leur condition auprès de leurs collègues (Drescher et Merlino, 2007). Rappelons que les DSM-I et II ne possédaient pas de définition formelle du trouble mental. Face à ces groupes de pression et sans définition explicite, l'APA n'était pas en mesure de justifier l'inclusion de l'homosexualité dans le manuel, ni son exclusion. Cette controverse a donc mis en évidence l'absence de balises claires pour savoir ce qui était compris comme un trouble

mental au sein du manuel. Face à cette controverse, la solution de Robert Spitzer, futur président de la *Task Force* du DSM-III, visait à proposer un compromis qui accommoderait les revendications des militants pour les droits des homosexuels et une position plus conservatrice tenue par plusieurs membres de l'APA, qui postulait que l'homosexualité était un trouble en soi. Ce compromis a été d'insister sur l'importance pour poser un diagnostic de trouble mental d'inclure la détresse comme critère diagnostique, faisant de l'homosexualité une condition irrégulière, mais pas pathologique en elle-même. Spitzer a donc présenté l'homosexualité comme une forme de comportement sexuel irrégulier, mais qui n'était pas pathologique en soi. L'homosexualité a donc été exclue du DSM en 1973, pour être remplacée par la catégorie « Ego-Dystonic Homosexuality » jusqu'en 1987 avec la parution du DSM-III-R. Pour cette nouvelle catégorie, la présence de détresse ou d'incapacité étaient alors les critères décisifs pour concevoir l'homosexualité comme un trouble mental. L'homosexualité était un trouble seulement si les homosexuels éprouvaient de la détresse face à leur orientation sexuelle. La catégorie a été définitivement retirée du DSM-III-R, non pas pour des raisons conceptuelles, mais parce qu'elle était peu utilisée dans la pratique clinique et que très peu d'articles avaient été publiés à son sujet (voir Kirk et Kutchins, 1992, p. 81-90).

Il est important de noter que cette controverse, bien que se manifestant à travers le cas de l'homosexualité, ciblait un enjeu conceptuel beaucoup plus général : celui de la définition du trouble mental. Comme l'affirment Kirk et Kutchins :

The debates about homosexuality could have been about most other diagnoses, had there been strong differences of opinions and hungry media. The debates had nothing to do with the ability of psychiatrists to identify homosexuals, but everything to do with a conceptual and theoretical problem, namely, whether homosexuality constituted a disorder. In order to address that question, psychiatrists would have to define disorder convincingly. (Kirk et Kutchins, 1992, p. 30)

De fait, le problème s'est présenté à nouveau face à d'autres conditions psychiatriques telles que les problèmes de fiabilité et de faux positifs du diagnostic de schizophrénie dans les années 1960 et 1970 (voir par exemple Garson, 2006, p. 56). Le cas de l'homosexualité est donc une manifestation d'un problème conceptuel plus général qu'est la définition du trouble mental, et auquel la psychiatrie a dû faire face à plusieurs reprises.

La question de la définition du trouble mental aujourd'hui

À ce stade de la présentation, un contre-argument doit être écarté : la définition du trouble mental n'est-elle pas qu'un problème historiquement situé, sans implication pour la psychiatrie contemporaine? Les critiques de l'antipsychiatrie et des groupes de pression pour les droits des homosexuels s'étant apaisées, la définition du trouble mental est-elle un enjeu aussi important aujourd'hui? La psychiatrie s'étant métamorphosée depuis les années 1950, le problème de la définition du trouble mental n'est pas disparu, mais se manifeste plutôt sous de nouvelles formes. Par exemple, on accuse aujourd'hui la psychiatrie de « surmédicalisation ». La « médicalisation » décrit le phénomène selon lequel des conditions jusqu'alors situées en dehors du champ médical sont définies et traitées comme des problèmes médicaux, en tant que « trouble » ou « maladie » (Conrad, 2007). La surmédicalisation est alors le fait de considérer certaines conditions comme faisant partie du domaine médical alors qu'elles ne devraient pas l'être (Conrad et Slodden, 2013). Ce phénomène a reçu une attention grandissante avec la publication du DSM-5 (American Psychiatric Association [APA], 2013) (p. ex. : Frances, 2010; Horwitz et Wakefield, 2012; Lane, 2007;). Il faut noter

que par contraste avec le mouvement antipsychiatrique, cet angle d'attaque ne remet pas en question l'existence des troubles mentaux, mais reproche plutôt à la psychiatrie l'étendue de son concept de trouble mental. Qui plus est, cette critique ne cible pas l'objectivité du concept de trouble mental, mais plutôt le fait que plusieurs états qui ne correspondent pas à ce qu'est un trouble mental se voient classés comme tels. Cet angle critique a été récemment défendu par Horwitz et Wakefield (Horwitz et Wakefield, 2005, 2007). Horwitz et Wakefield se concentrent principalement sur le diagnostic de dépression. Leur critique repose sur le caractère trop inclusif de la définition de trouble du DSM, ce qui aurait pour effet que des individus sains, mais souffrant d'une tristesse normale causée par des événements difficiles (p. ex. : le deuil), seraient diagnostiqués comme dépressifs. Cet angle d'attaque repose sur l'idée qu'il existe un continuum sur lequel se situent les troubles mentaux et les expériences normales de vie. La psychiatrie aurait en ce sens mal identifié le seuil à partir duquel on peut désigner un état comme pathologique. Wakefield et Horwitz défendent, suite à cette critique, que la psychiatrie devrait restreindre l'étendue du concept de trouble mental à des déviations par rapport aux normes de fonctionnement évolutives de l'espèce. En ce sens, ces auteurs reprochent à la psychiatrie de ne pas prendre au sérieux la définition « réelle » du trouble mental.

Le mouvement récent de la neurodiversité est également lié à l'enjeu conceptuel de la définition du trouble mental (voir Armstrong, 2010; Chapman, 2020; Silberman, 2015). Ce phénomène est apparu à la fin des années 1990, principalement au sein des communautés de personnes autistes voulant que la psychiatrie (et la société) les reconnaisse comme des personnes différentes, mais pas dysfonctionnelles³. Aujourd'hui, ce mouvement a pris de l'ampleur, avec la création de groupes de soutien

³ Des parallèles entre le mouvement de la neurodiversité et les revendications de certaines personnes sourdes militantes peuvent être tracés (voir Cooper, 2007, 2021). Ces militants soutiennent que le fait d'être sourd n'est pas un handicap ou une dysfonction, mais une simple différence vis-à-vis la norme dominante, celle des personnes capables d'entendre.

ainsi que l'apparition de nombreux sites internet et blogues sur le sujet (p. ex. : le Developmental Adult Neuro-Diversity Association (DANDA), les sites www.neurodiversity.com et www.neurocosmopolitanism.com, etc.). Les revendications de ce mouvement se sont désormais généralisées à plusieurs conditions considérées comme des troubles mentaux par le DSM, principalement les troubles neurodéveloppementaux (tels que le spectre de l'autisme, le trouble du déficit de l'attention avec hyperactivité (TDAH), la dyslexie, le syndrome de Tourette, etc.). Le mouvement de la neurodiversité plaide pour que ces conditions soient reconnues comme des variations (neurophysiologiques) non pathologiques. Dans ce cadre, on parle de personne neurotypique pour référer aux individus dont le fonctionnement neurophysiologique est statistiquement normal, et au de neuro-atypique pour désigner les individus dont le fonctionnement est différent de cette norme. Les individus neuro-atypiques présenteraient donc seulement des différences par rapport aux individus neurotypiques, mais leur différence ne devrait pas être perçue comme pathologique. La revendication politique que porte ce mouvement est celle de la reconnaissance de formes alternatives de fonctionnement neurophysiologique, et leur dé-pathologisation. À la différence de la critique précédente, qui prend appui sur une définition en partie biomédicale du trouble mental (en termes de dysfonction interne à l'individu), le mouvement de la neurodiversité remet en question le paradigme biomédical pour conceptualiser de telles conditions, du moins pour penser les troubles neurodéveloppementaux (Chapman, 2020, p. 373-375). Pour les défenseurs du mouvement, les normes qui guident la distinction entre le normal et le pathologique quant aux troubles neurodéveloppementaux sont uniquement sociales. La psychiatrie devrait donc viser à améliorer la qualité de vie des personnes neuro-atypiques, sans viser à les « normaliser » pour les faire correspondre à cette norme sociale. Le mouvement de la neurodiversité est controversé, à la fois en psychiatrie, mais aussi au sein des populations de personnes présentant des troubles neurodéveloppementaux (p. ex. : Chapman, 2020; Jaarsma et Welin, 2012). Certaines critiques du mouvement

pointent l'invisibilisation potentielle de la souffrance et de la détresse de plusieurs personnes étiquetées comme neuroatypiques par le mouvement (p. ex. : les personnes autistes qui ne sont pas hautement fonctionnelles). Évaluer les revendications du mouvement de la neurodiversité demande que l'on se positionne sur ce qu'est un trouble mental. Autrement dit, répondre à la question « les troubles neurodéveloppementaux sont-ils pathologiques, ou des formes alternatives de fonctionnement? » dépend de la manière dont on comprend les concepts de « pathologique » et de « troubles mentaux » (Shields et Beversdorf, 2020).

Ainsi, même si l'approche adoptée dans cette introduction est davantage historique, il faut insister sur l'importance de la question de la définition du trouble mental dans les débats contemporains visant la psychiatrie. Il faut donc garder en tête que même si le débat sur le concept de trouble mental trouve son origine dans les années 1950, il s'agit d'un enjeu de taille qui mérite une attention renouvelée plutôt que purement historique.

Les définitions formelles du désordre mental à partir du DSM-III

Revenons au tournant des années 1970 et aux différents angles d'attaque initiaux qui ont poussé la psychiatrie à réviser la deuxième édition du DSM et à adopter le modèle biomédical, dans le but de réaffirmer sa légitimité comme sous-discipline de la médecine (Guze, 1992; Murphy, 2005; voir Feighner *et al.*, 1972 et Klerman, 1978 pour une défense du modèle médical). Cette transformation a eu pour conséquence, entre autres, le fait que la recherche et les traitements devaient désormais correspondre au raisonnement et à la méthode médicale. Qui plus est, les troubles mentaux devaient

être conçus comme des pathologies médicales, c'est-à-dire comme des dysfonctions internes des parties de l'organisme. Le souhait était que la classification des conditions psychiatriques en tant que troubles mentaux soit guidée par des faits objectifs, et non pas par des valeurs (du moins en partie selon Wakefield, 1992; voir aussi Guze, 1992). Conformément à ce dernier point, le passage du modèle psychanalytique au modèle biomédical illustre l'importance d'adresser l'enjeu de la définition du trouble mental en vue de légitimer la scientificité et l'objectivité de la psychiatrie (Porter, 2002, chapitre 8; voir Spitzer et Endicott, 1978 pour l'insertion de la notion de dysfonction comme conforme au modèle médical, mais aussi Demazeux, 2021 sur le rôle limité de cette notion pour les deux auteurs). C'est pourquoi, dans cette section, je propose d'examiner de manière critique les différentes définitions formelles du DSM.

En 1973, l'APA a mis sur pied une équipe de travail ayant pour mandat d'élaborer une définition formelle du trouble mental, soit la *Task Force on Nomenclature and Statistics* (Kirk et Kutchins, 1992, cette équipe de travail avait aussi pour mission de définir le terme « psychiatre »). À la tête de cette équipe de travail était Robert Spitzer, nommé comme président, en partie en raison de ses habiletés à calmer les tensions concernant l'inclusion de l'homosexualité au sein du DSM (voir section précédente). La création de cette équipe de travail est la manifestation de la croyance de plusieurs psychiatres membres de l'APA pour qui l'élaboration d'un ensemble de critères définissant le trouble mental était ce qui pourrait résoudre la crise⁴. En effet, la définition objective du trouble mental était vue par plusieurs comme ce qui pourrait distinguer formellement les troubles mentaux des états psychologiques ou des comportements jugés déviants par la société dans le cadre d'un modèle biomédical de la psychiatrie (Cooper, 2005, 2015; Spitzer et Endicott, 1978).

⁴ Il est à noter que cette équipe de travail n'a cependant pas produit de définition du désordre mental, mais plutôt un dépliant intitulé *What is a psychiatrist?*, que les cliniciens devaient distribuer à leurs patients (voir Cooper, 2015).

C'est lors de la parution du DSM-III (APA, 1980) que la formulation de la première définition officielle de la catégorie générale de trouble mental a été publiée. Il est à noter cependant que l'APA reconnaissait déjà l'absence de frontières strictes qui délimiteraient avec précision le phénomène de trouble mental. La définition se veut donc une caractérisation visant à expliquer les critères qui justifient généralement l'inclusion ou l'exclusion de certaines catégories diagnostiques plutôt qu'une définition « absolue » du trouble mental (APA, 1980; voir Garson, 2006 pour une discussion). La formulation du DSM-III présente le trouble mental de manière partiellement objective comme une dysfonction interne présente chez l'individu, conceptuellement distincte des préjudices occasionnés par cette dysfonction :

In DSM-III each of the mental disorders is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is typically associated with either a painful symptom (distress) or impairment in one or more important areas of functioning (disability). In addition, there is an inference that there is a behavioral, or biological dysfunction, and that the disturbance is not only in the relationship between an individual and society. (When the disturbance is limited to a conflict between an individual and society, this may represent social deviance, which may or may not be commendable, but is not by itself a mental disorder) (APA, 1980).

La motivation derrière cette définition était donc d'exclure les comportements et états psychologiques « problématiques », mais non pathologiques. Dans la définition, cette exclusion repose sur l'insertion de deux critères, soit la dysfonction interne et la détresse ou le handicap (voir Spitzer et Endicott, 1978 pour une discussion des arguments qui ont motivé l'intégration de ces différents critères). Ainsi, avec cette définition, le manque de talent ou la témérité ne peuvent être considérés comme des troubles, tout comme la déviance sociale, puisqu'ils n'impliquent pas de dysfonction interne (Wakefield et First, 2003). Qui plus est, il n'est pas suffisant qu'une dysfonction soit présente pour que l'on puisse désigner l'état comme pathologique, étant donné que

cette dysfonction peut ne pas avoir suffisamment de conséquences négatives sur le bien-être de l'individu. C'est donc seulement lorsque des expériences de détresse ou de handicap significatives reposent sur une dysfonction interne sous-jacente à ces formes de préjudices que l'on peut parler de trouble mental.

Bien que cette définition ait eu des effets positifs sur le développement de la psychiatrie par l'importance qu'elle accorde au critère de détresse ou de handicap pour poser un diagnostic (Wakefield et First, 2003), il n'est pas clair qu'elle répondait bien aux critiques mises en lumière dans la section précédente. En effet, malgré le fait que la définition ait pour objectif d'atténuer l'exclusion ou la coercition sociale causées par les diagnostics psychiatriques, l'objectivité du concept de trouble mental n'en a pas été pour autant fondée. En effet, la définition tente de séparer la dysfonction et la déviance sociale de manière négative, en posant que la dysfonction ne doit pas seulement être le reflet de normes sociales. Or la définition n'explique pas ce qu'est une dysfonction, ni comment cette dysfonction peut être dite objective (Garson, 2006; Wakefield et First, 2003) (je reviens sur cet enjeu plus loin). Qui plus est, le fait de mettre de l'avant les critères de détresse ou de handicap comme conditions nécessaires au diagnostic diffère de l'approche biomédicale. Du point de vue médical, l'état d'un individu peut être qualifié de pathologique s'il s'agit d'une dysfonction interne, même si cet individu ne ressent aucune détresse ni incapacité. Ce critère est donc en tension avec l'objectif de rendre le concept de désordre mental équivalent à celui de désordre physique (Cooper, 2015; Varga, 2015, p. 114).

Bien que la définition des DSM-III et du DSM-III-R ne répondait pas complètement aux critiques ayant motivé sa formulation, la quatrième édition du DSM ne l'a que peu modifiée. Cependant, certaines précisions ont été ajoutées quant au statut de cette définition. Ainsi, dans le DSM-IV, on peut lire :

Moreover, although this manual provides a classification of mental disorders, it must be admitted that no definition adequately specifies precise boundaries for the concept of “mental disorder”. The concept of mental disorder, like many other concepts in medicine and sciences, lacks a consistent operational definition that covers all situations. All medical conditions are defined on various levels of abstraction – for example, structural pathology (e.g. ulcerative colitis), symptom presentation (e.g. migraine), deviance from a physiological norm (e.g. hypertension), and etiology (e.g. pneumococcal pneumonia). Mental disorders have also been defined by a variety of concepts (e.g. distress, dyscontrol, disadvantage, disability, inflexibility, irrationality, syndromal pattern, etiology, and statistical deviation). Each is a useful indicator for a mental disorder, but none is equivalent to the concept, and different situations call for different definitions.

Despite these caveats, the definition of mental disorder that was included in DSM-III and DSM-III-R is presented here because it is as useful as any other available definition and has helped to guide decisions regarding which conditions on the boundary between normality and pathology should be included in DSM-IV. In DSM-IV, each of the mental disorders is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in the individual and that is associated with present distress (e.g. a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) or with a significant increased risk of suffering death, pain, disability, or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one. Whatever its original cause, it must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual. Neither deviant behavior (e.g. political, religious, or sexual, nor conflicts that are primarily between the individual and society) are mental disorders unless the deviance of conflict is a symptom of a dysfunction in the individual, as described above (APA, 1994, les italiques sont les miennes).

Le DSM-IV persiste donc avec une définition du désordre mental très semblable à celle des DSM-III et des DSM-III-R dans son contenu, en ce qu'elle conserve les critères de détresse et de handicap ainsi que l'idée de dysfonction sous-jacente. Deux variations peuvent cependant être notées. D'abord, comme le souligne Cooper (2015), dans le DSM-IV et le DSM-IV-R, la notion de détresse est davantage perçue comme ce qui pourrait déterminer le *seuil* à partir duquel un état est considéré comme pathologique. Dans le DSM-III et le DSM-III-R, le critère de détresse vise plutôt l'inclusion ou

l'exclusion de catégories diagnostiques spécifiques au sein du manuel (comme l'exclusion de l'homosexualité). Ce qui diffère véritablement entre les DSM-III et IV, c'est la reconnaissance du flou entourant le concept trouble mental. Dans le DSM-IV, l'incapacité de la définition du trouble mental à pouvoir découper l'ensemble des conditions psychiatriques et de le distinguer du normal est reconnue de manière explicite⁵.

En ce qui concerne la parution de la cinquième édition du DSM, face à cette absence de critères clairs quant à ce qui devrait distinguer le trouble mental de l'état sain, le groupe de travail sur la nomenclature de l'APA a nommé comme l'une de ses priorités la formulation d'une définition du désordre mental qui parviendrait à établir cette frontière (Rounsaville *et al.*, 2002) : « If for no other reason, this is important because of rising public concern about what is sometimes seen as the progressive medicalization of all problem behaviors and relationships. » C'est donc dire que la définition du trouble mental était toujours perçue, lors de la mise en place du processus de révision du DSM-5, comme une tâche centrale. Notons par ailleurs que les auteurs du groupe de travail reconnaissent qu'une définition claire pourrait apaiser les controverses contemporaines entourant la médicalisation (et la surmédicalisation) en psychiatrie. Dans le DSM-5, le trouble mental est défini comme suit :

A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually

⁵ Une hypothèse pour ce changement de ton est le changement à la présidence du comité de révision du DSM. Alors que les DSM-III et les DSM-III-R ont été dirigés par R. Spitzer, qui reconnaissait pleinement l'utilité d'une définition pour régir l'inclusion ou l'exclusion des conditions dans le manuel, Allen Frances rejetait la potentielle utilité d'une telle définition. Qui plus est, la parution du DSM-IV et du DSM-IV-R s'est déroulé dans un contexte plus calme politiquement. Le besoin de clarifications conceptuelles était beaucoup moins criant que dans les années 1960 et 1970 (voir Cooper, 2015 à ce sujet).

associated with significant distress or disability in social, occupational, or other important activities. An expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above. (APA, 2013, p. 20)

Ici, le statut du critère de détresse est beaucoup moins clair que dans les formulations précédentes, étant donné que la définition pose seulement que la détresse est « généralement associée » au désordre (Cooper, 2015). Il ne s'agit donc pas d'une condition nécessaire, mais plutôt d'une cooccurrence. Le critère de déviation par rapport à une norme de fonctionnement mental demeure quant à lui à peu près inchangé. Conformément aux formulations précédentes, le DSM-5 affirme qu'il n'est pas suffisant pour un individu d'éprouver de la détresse ou d'être le sujet d'incapacité en lien avec sa condition. Par ailleurs, comme dans les éditions passées, l'accent mis sur la dysfonction interne implique aussi que la simple déviance sociale n'est pas suffisante pour pouvoir poser le diagnostic de trouble. C'est donc la présence d'une dysfonction interne qui devrait fondamentalement permettre d'identifier si un état est pathologique ou non.

Le DSM s'est donc doté d'une définition du trouble mental à partir de sa troisième édition, dans un contexte de crise de légitimité, et cette définition est demeurée à peu près la même au fil des révisions du DSM, jusqu'au DSM-5. Du point de vue de la description et de la classification des conditions psychiatriques spécifiques (objectifs 1 et 2 de Bolton, voir section précédente), l'avènement du DSM-III et la parution des éditions suivantes ont fait progresser la psychiatrie dans sa quête d'objectivité et de scientificité. Le passage du cadre psychanalytique vers un cadre observationnel et descriptif propre au modèle médical a augmenté la fiabilité des diagnostics particuliers et la facilité de communication entre les différents acteurs impliqués dans le système de santé mentale, principalement les cliniciens. L'approche descriptive des symptômes

et la classification en syndromes, avec son caractère a-théorique semble avoir permis de se rapprocher des exigences d'objectivité de la médecine. Le DSM est désormais l'outil de communication principal entre les différents acteurs de la pratique clinique en psychiatrie, mais aussi des systèmes administratifs et légaux. C'est aussi l'outil qui guide, pour le moment, la recherche clinique (voir cependant Hyman, 2010; Murphy, 2006; Tabb, 2015; Tsou, 2015 pour des critiques de la validité des catégories actuelles; pour une discussion sur le *Research Domain Criteria* (RDoC), un cadre de recherche alternatif, voir Cuthbert, 2015).

Il n'est toutefois pas évident que le même type de progrès en termes de scientificité ait été atteint par la définition actuelle du trouble mental. Comme l'explique bien Bolton :

Interrater reliability simply does not settle the issue one way or another. It is settled, rather, by inspecting the content of the descriptions, and what we find in the case of psychiatric symptoms and syndromes is pervasive use of norms and, perhaps, values. This implies that what we are all agreeing about, what we can recognize as if by eye alone, turns on a more or less implicit theory as to what is and is not normal mental and behavioral functioning. But what is this implicit theory, and what is its basis? (Bolton, 2008, 15)

L'atteinte d'une certaine fiabilité et d'une validité des diagnostics spécifiques de troubles mentaux ne résout pas le problème de la validité de la catégorie générale de trouble mental. En effet, déjà lors de la première formulation du DSM-III, il ne semblait pas que cette définition puisse répondre aux critiques des antipsychiatres et de la justification de l'inclusion ou de l'exclusion de l'homosexualité. La définition, plutôt que de se pencher sur ce qui pourrait fonder objectivement le concept de trouble mental, décrit ce qui est traité par les psychiatres. Comme plusieurs philosophes l'ont noté, la principale difficulté pour définir objectivement le concept de trouble mental réside dans le caractère intrinsèquement normatif de ce concept, problème que n'affronte pas la définition (voir Bolton, 2001; Garson, 2006). Il semble donc y avoir une tension entre le désir d'objectivité et de scientificité de la psychiatrie et le fait que l'un de ses

concepts centraux, le trouble mental, soit de manière inhérente normatif. L'ajout du critère de dysfonction interne était ce qui devait assurer l'objectivité du concept. Malgré les différentes propositions de définition du DSM, le problème de l'objectivité et de la scientificité du concept de trouble mental demeure entier, puisque les définitions ne parviennent pas à établir si cette norme de fonctionnement mental impliquée dans la définition est le reflet de normes sociales et culturelles ou provient de faits scientifiques. Pour répondre aux critiques des antipsychiatres et de la médicalisation des problèmes normaux de vie, il a paru pour plusieurs que la définition devrait pouvoir montrer que le concept réfère à des entités naturelles qui ne dépendent pas uniquement des normes sociales.

Ainsi, au sein des définitions du DSM, l'objectivité du concept de *dysfonction* est un angle mort. Le principal problème est que le DSM a la volonté de demeurer a-théorique et descriptif en ce qui relève des causes ou des déterminants des symptômes constitutifs des désordres (voir par exemple Wakefield, 1992a). La définition demeure donc muette quant à ce qui pourrait réellement assurer la neutralité du concept de dysfonction. Or, rappelons que le but de l'ajout de cette composante à la définition était justement d'assurer un fondement objectif à la catégorie générale de désordre mental. La dysfonction est cependant définie de manière circulaire (Fulford, 2000; Garson, 2006). En effet, la dysfonction est définie par rapport aux symptômes cliniques qui lui sont associés, ce que ce concept devait pourtant expliquer. En offrant une définition neutre du point de vue théorique, conformément au projet de neutralité du manuel, il est évident que la définition ne peut répondre aux critiques visant le caractère médico-scientifique du concept. Ainsi, *au sein du DSM*, les normes implicites au concept de trouble mental, comme présentées dans sa définition, ne sont pas fondées sur des faits scientifiques, mais reposent plutôt sur une circularité problématique. Sortir de cette circularité impliquerait de fournir des fondements ou des justifications objectifs qui rendraient compte de la norme de fonctionnement supposée par la définition (voir

Garson, 2006; Wakefield, 1992a). Qui plus est, il est à noter que lors de la révision du DSM-5, aucun groupe de travail n'a finalement été formé pour se pencher sur une proposition de définition révisée du trouble mental. Un travail d'analyse et de justification reste donc à faire pour que la définition du DSM soit pleinement satisfaisante. Ce sont ces enjeux qui ont occupé une bonne partie de la recherche en philosophie de la médecine et de la psychiatrie, traités principalement par l'approche objectiviste des fonctions naturelles.

La question de la définition du trouble mental sous l'angle philosophique : deux grandes approches

Parallèlement aux discussions impliquant les psychiatres et leurs opposants, les philosophes se sont également emparés de la question de la définition du trouble mental dès le début de ces controverses. Or, la manière dont la question du trouble mental a été traitée par les philosophes n'est pas exactement la même que celle des psychiatres. Comme le notent à juste titre Aucouturier et Demazeux (2013, p. 86), le problème de la définition du trouble mental a pris une tournure conceptuelle et est devenu beaucoup plus analytique lorsqu'appréhendé sous l'angle philosophique. De plus, la distinction entre faits et valeurs est devenue une dimension centrale du débat. De ce point de vue, depuis les années 1970, les définitions proposées par les philosophes en réponse à la question « qu'est-ce que le trouble mental? » se sont inscrites sur un continuum allant d'un concept totalement indépendant des valeurs [*value-free*] à un concept exclusivement influencé par des valeurs sociales et culturelles [*value-laden*]. Le débat philosophique sur la définition du trouble mental implique une vaste documentation et

se caractérise par de nombreuses approches concurrentes, que cette introduction ne me permet pas de couvrir de manière exhaustive (voir Cooper, 2011 pour une revue plus complète). C'est pourquoi je me consacrerai principalement à deux grandes approches qui tentent de fournir une réponse à la question des valeurs dans la définition du trouble mental. La première est appelée « objectivisme des fonctions naturelles » et la seconde, « constructivisme » ou « normativisme ».

Rappelons que le travail d'élaboration d'une définition du trouble mental qui pourrait accommoder les critiques de la psychiatrie supposait que si le trouble mental pouvait être défini en des termes objectifs et scientifiques, alors la psychiatrie pourrait légitimer son statut de branche de la médecine, et répondre ainsi à ses critiques. En effet, avec une définition objective, les cliniciens pourraient distinguer le trouble mental de la déviance sociale et des réactions normales face aux aléas de l'environnement, répondant respectivement aux critiques des antipsychiatres et de la médicalisation des problèmes normaux de vie, en plus de fournir un critère pour départager des catégories diagnostiques controversées. Une approche naturaliste et scientifique pour définir le concept de désordre mental a donc semblé à plusieurs une réponse prometteuse à ces attaques (Woolfolk, 1999). Parallèlement aux différentes définitions du DSM qui se veulent agnostiques quant à la nature du concept de dysfonction, les philosophes de la médecine et de la psychiatrie ont proposé des tentatives de définition objective du trouble en se fondant sur l'élaboration de la notion de dysfonction. L'approche objective la plus connue est celle de l'objectivisme des fonctions naturelles [*natural function objectivism*] (j'emprunte ici l'appellation proposée par Varga, 2011, 2015). Selon cette approche, il est possible de définir le trouble mental comme une déviation par rapport au fonctionnement normal relativement à un groupe de référence donné. Cette approche réfère à l'idée selon laquelle le concept de trouble mental est objectif en ce qu'il ne réfère pas à des valeurs, mais seulement à des normes de fonctionnement naturelles. Au sein de cette approche, on peut distinguer les contributions de

Christopher Boorse (p. ex. : Boorse, 1976, 1977, 1997, 2014) et de sa théorie biostatistique de la santé (BST), ainsi que de Jerome Wakefield (p. ex. : Wakefield 1992b, 1999) et de sa théorie de la dysfonction préjudiciable (DP)⁶.

L'approche objectiviste des fonctions naturelles défendue par Boorse et Wakefield suppose que le trouble réfère à une dysfonction interne à l'organisme par rapport à un standard de fonctionnement physiologique. Les deux théories ont été largement discutées et critiquées au sein des écrits philosophiques. En effet, plusieurs ont remis en question la capacité de telles approches de fonder l'objectivité du concept de trouble en des termes purement physiologiques et objectifs, angle critique qui a visé Boorse principalement (contre Boorse, voir par exemple Agich, 1983; Cooper, 2005; DeVito, 2000; Ereshefsky, 2009; Fulford, 2001; Kingma, 2007, 2010, 2013; Varga 2011; contre Wakefield, voir Kirmayer et Young, 1999; Sadler et Agich, 1995). Wakefield a quant à lui été abondamment critiqué pour son concept étiologique de dysfonction (voir Bolton, 2008; Kingma, 2013) et pour l'aspect révisionniste de son approche (Lemoine, 2021; Murphy et Woolfolk, 2000b). Bien qu'en apparence prometteuse, il semble désormais que l'approche objectiviste des fonctions naturelles ne soit pas en mesure d'assurer l'objectivité du concept de dysfonction, et par là ne pourrait pas répondre aux critiques du trouble mental utilisé en psychiatrie.

Par opposition, les normativistes définissent le plus souvent le trouble mental en référence à des conventions sociales et culturelles. Cette définition impliquerait alors des jugements évaluatifs, et serait donc influencée par des valeurs sociales et culturelles. Selon Rachel Cooper (2002, 2005), qui représente bien le courant du normativisme, la santé et la maladie seraient des concepts de nature anthropocentrique,

⁶ D'autres conceptions objectivistes des fonctions naturelles ont été proposées par le passé, voir par exemple Kendell 1975; Scadding 1990; Tyrer et Steinberg 2003. Or, Boorse et Wakefield sont les deux auteurs qui ont eu le plus d'influence au sein de cette approche et qui ont proposé les approches les plus élaborées, ce pour quoi je les discute principalement.

qui reflètent ce qui est d'intérêt pour les êtres humains. Au terme de son analyse, Cooper propose trois critères conjoints pour définir la pathologie à la fois du point de vue mental et physique. Cooper s'écarte ici de l'approche des fonctions naturelles en ce qu'elle ne pose pas comme critère définitionnel la dysfonction, mais insiste plutôt sur le caractère normatif et pratique de la pathologie. L'intérêt de l'approche normativiste est de rendre compte de nos intuitions selon lesquelles un concept théorique de pathologie, qu'il concerne le mental ou le physique, ne peut être détaché de l'aspect pratique de la médecine. La primauté est ainsi accordée au jugement de valeur qui permet d'identifier un état comme pathologique. Or l'approche normativiste présente également certains problèmes caractéristiques, le principal étant que la reconnaissance des valeurs sociales et culturelles qui influencent le concept de trouble mental semble limiter les prétentions scientifiques et objectives de la psychiatrie.

Depuis les années 1970, le débat portant sur la définition de la santé mentale a principalement consisté en une évaluation de chacune de ces deux approches. Or plusieurs auteurs ont récemment qualifié la manière dont le débat s'était développé de « cul-de-sac » (voir par exemple Murphy, 2005; Schuur, 2020). En effet, au regard de l'évolution du débat, une première conclusion émerge : les deux pôles que sont l'approche objectiviste des fonctions naturelles et le normativisme sont inadéquats pour rendre compte de la complexité du phénomène qu'est le trouble mental. Il semble que la signification que l'on accorde au trouble mental contienne à la fois des valeurs sociales et culturelles et des faits scientifiques. Qui plus est, la notion de dysfonction ne semble pas être parfaitement indépendante des valeurs. Cette reconnaissance a plus récemment conduit à des vues nuancées dans lesquelles les valeurs sociales et culturelles et les faits scientifiques sont inclus, tant en philosophie de la médecine qu'en philosophie de la psychiatrie. Deux propositions me semblent particulièrement intéressantes parce qu'elles proposent de recadrer le débat dans un cadre épistémologique particulier (De Vreese, 2017; Kingma, 2014; voir Varga, 2017 pour

une discussion sur d'autres approches). D'abord, Kingma (2014) propose d'adopter l'approche du constructivisme social. Selon l'interprétation de Kingma, le constructivisme social conçoit les construits théoriques comme le résultat de processus sociaux et culturels. En ce sens, la signification de ces construits est contingente et ne peut être comprise que par rapport aux structures sociales et culturelles dans lesquelles elle a été formée. Dans le cadre du constructivisme social, une classification peut refléter certains aspects sociaux ou culturels, tout en étant décrite de manière naturaliste et en des termes empiriques. Pour Kingma, l'intérêt du constructivisme social réside dans sa capacité à accommoder les points forts de l'approche normativiste et de l'objectivisme des fonctions naturelles. L'approche normativiste aurait bien mis en évidence les considérations évaluatives qui se rapportent à notre concept de santé, tandis que l'approche objectiviste aurait insisté sur les aspects scientifiques de la pathologie comme une dysfonction. Selon Kingma, le constructivisme social permettrait de rendre compte de ces deux intuitions. Dans une optique similaire, De Vreese (2017) a récemment proposé de recadrer le débat au sein de l'approche pragmatiste. De Vreese défend que notre conception de la santé est teintée à la fois de considérations qui proviennent de faits scientifiques et de valeurs. Les valeurs réfèrent à des états psychologiques ou physiologiques jugés indésirables, tandis que les faits scientifiques réfèrent aux causes qui permettent d'expliquer ces états dans une perspective d'intervention thérapeutique⁷. De ce point de vue, la connaissance scientifique de la médecine concerne les processus causaux pertinents dans l'explication des états jugés indésirables. Ces deux propositions me semblent particulièrement intéressantes parce qu'elles permettent de dépasser l'opposition mutuellement exclusive entre normativisme et approche objectiviste des fonctions naturelles, et qu'elles maintiennent le rôle des faits scientifiques dans leur conception de la pathologie malgré l'intégration d'éléments axiologiques. Or jusqu'à maintenant, ces deux approches sont au stade de « propositions de recherche » et aucune ne permet

⁷ Cette position est semblable à celle défendue par Engelhardt (1975).

de répondre véritablement aux critiques de l'antipsychiatrie ou de la médicalisation des problèmes normaux de vie. En effet, il semble qu'une tâche essentielle lorsque l'on admet la présence de valeurs au sein du concept de trouble mental est de montrer que ces valeurs n'empêchent pas la psychiatrie d'accéder à son titre de science. Autrement dit, il faut démontrer que reconnaître que les valeurs influencent notre compréhension de ce qu'est le trouble mental n'implique pas de donner raison aux normativistes radicaux, qui nieraient une quelconque base scientifique et objective au concept. Recadrer le débat sur la définition du trouble mental devrait permettre de répondre à ce défi de taille. Pour l'instant, ni l'approche de Kingma, ni l'approche de De Vreese n'offrent une solution pleinement satisfaisante à ce problème.

Jusqu'à maintenant, les propositions qui tentent d'intégrer faits et valeurs à la définition demeurent au stade de programme de recherche. La manière exacte dont les faits et les valeurs interagissent dans le cas du trouble mental demande donc encore à être clarifiée. Un concept de trouble mental à la fois social et objectif reste encore à être intégré dans un cadre épistémologique solide. Selon moi, une approche qui fournit des ressources conceptuelles pour penser ce problème est celle de l'épistémologie sociale des sciences. L'épistémologie sociale se présente de manière générale comme une alternative à l'épistémologie traditionnelle, qui s'intéresse aux croyances individuelles des individus et aux critères qui permettent d'évaluer si un individu a raison d'avoir certaines croyances. L'épistémologie traditionnelle s'est intéressée à la production et à l'évaluation de connaissances individuelles, indépendamment de l'influence épistémique que peut avoir la situation sociale dans laquelle se trouvaient les individus. Par contraste, l'épistémologie sociale a pour but de mettre à l'avant-plan l'impact épistémique des interactions et des institutions sociales sur la production et l'évaluation des connaissances (p. ex. : désaccord entre les pairs, évaluation d'un témoignage, croyances collectives, conséquences épistémiques de l'organisation sociale, etc., voir Goldman et Blanchard, 2018). L'épistémologie sociale *des sciences* s'intéresse aux

aspects sociaux de la science, qu'il s'agisse des interactions sociales ou des institutions sociales, et à la manière dont les différentes dimensions sociales de l'investigation scientifique influencent la production de connaissances scientifiques. Influencée par la sociologie des sciences et l'histoire des sciences, l'épistémologie sociale des sciences se distingue pourtant de ces deux disciplines par le caractère normatif de son analyse de la science : elle ne cherche pas seulement à décrire les processus sociaux qui mènent à la production de connaissances scientifiques, mais aussi à fournir des normes épistémiques guidant cette production (voir Fuller, 1988; Schmitt, 1994). Se tourner vers l'épistémologie sociale des sciences pour penser le problème des valeurs sociales associées au concept de trouble mental me semble nécessaire pour éclairer le double caractère scientifique et normatif du trouble mental.

De récents arguments qui soutiennent l'utilisation des outils conceptuels de l'épistémologie sociale pour penser plusieurs problèmes en psychiatrie ont été récemment proposés. Par exemple, Anke Bueter a défendu la pertinence de lier les écrits en épistémologie sociale à ceux en philosophie de la psychiatrie (bien que comme elle l'admette, cette liaison en est à ses balbutiements) (Bueter, 2019; voir aussi Ross, 2002 en psychiatrie et Kincaid, 2007 en philosophie de la médecine). Parce qu'elle s'intéresse aux dimensions sociales de la production et l'évaluation de la connaissance, l'épistémologie sociale semble appropriée pour traiter plusieurs problèmes en philosophie de la psychiatrie. Selon Bueter (2019), cette pertinence est soutenue par trois raisons principales : 1) la psychiatrie, de par sa nature pratique, a un effet sur les représentations sociales de la maladie mentale et de l'organisation des soins de santé; 2) la psychiatrie est affectée par les structures sociales au sein desquelles elle se développe, qu'il s'agisse d'élaborer les critères de succès épistémique, les présuppositions qui guident la pratique et la recherche psychiatrique ou l'influence du financement sur les questions de recherche poursuivies; 3) le statut controversé de la psychiatrie en tant que science dans nos sociétés occidentales contemporaines. Des

thèmes clés de l'épistémologie sociale, tels que le témoignage, l'expertise ou l'agentivité épistémique, trouvent alors un écho intéressant lorsque traités en psychiatrie. Plus propres à ma thèse sont les thèmes traités par l'épistémologie sociale des sciences.

L'épistémologie sociale des sciences, parce qu'elle s'intéresse aux aspects sociaux qui peuvent influencer positivement ou négativement le succès épistémique d'une discipline scientifique, est un outil prometteur pour la philosophie de la psychiatrie. La recherche pharmaceutique et l'expertise médicale, l'influence des valeurs en psychiatrie et la place des patients dans le processus de révision du DSM sont des objets d'études tout à propos pour l'épistémologie sociale des sciences, lorsqu'appliquée à la psychiatrie. Comme l'affirme Bueter (2019), l'objectivité scientifique ainsi que la place et le rôle des valeurs en science sont des thèmes particulièrement pertinents pour la philosophie de la psychiatrie. Plus spécifiquement, plusieurs auteurs en épistémologie sociale des sciences ont remis en question la thèse de l'idéal de la science indépendante des valeurs, selon laquelle les valeurs non épistémiques ne devaient pas entrer dans les phases internes de la science (les phases de justification épistémique, c'est-à-dire d'évaluation des théories ou hypothèse par rapport aux données empiriques). Cet idéal a été remis en question sous plusieurs angles (p. ex. : Douglas, 2000, 2009; Dupré, 2007; Elliott 2011, p. 55-80, Longino, 1990, 2002; Putnam, 2002; Wilholt, 2009). Tous ces écrits trouvent leur pertinence en philosophie de la psychiatrie, étant donné que l'influence des valeurs sociales et culturelles sur la recherche et la pratique psychiatrique y a été souvent mise en évidence et critiquée. Plus en lien avec le thème de cette thèse, le débat qui concerne la définition du trouble mental, parce qu'il concerne l'influence ou l'absence d'influence des valeurs sur cette définition, pourrait tirer profit d'une mise en relation avec les outils de l'épistémologie sociale des sciences.

J'aimerais ajouter à l'analyse de Bueter, qui a pour but de dresser un portrait général des liens prometteurs entre philosophie de la psychiatrie et épistémologie sociale des sciences, en insistant sur la manière dont l'approche plus spécifique d'Helen Longino (1990, 2002) peut permettre de répondre au problème de la définition du trouble mental. La conception de *l'objectivité sociale* que Longino élabore met de l'avant un concept d'objectivité différent de celui que l'on retrouve traditionnellement dans les écrits, plus précisément un concept d'objectivité qui prend en compte et intègre l'influence des valeurs sociales et culturelles en science. Dans cette optique, l'objectivité dépend des processus de production du savoir scientifique, qui ont pour but de rendre explicite et de critiquer les valeurs sociales et culturelles présentes dans le raisonnement scientifique.

Comme mis en lumière précédemment, les questionnements philosophiques sur l'objectivité du concept de trouble mental sont nés dans un contexte de critiques qui remettaient en question la légitimité scientifique du concept. Le débat autour de l'homosexualité, conçu comme un trouble mental dans les DSM I et II, est l'illustration typique de l'intrusion de biais défavorisant les groupes marginalisés au sein de la classification psychiatrique. Le concept d'objectivité sociale défendue par Longino permet d'éclairer ce type de dynamiques et de fournir une réponse pertinente et appropriée aux critiques de la définition. En effet, l'approche de Longino s'intéresse aux dynamiques de pouvoir et aux relations d'oppression qui peuvent influencer la production de connaissances scientifiques, ainsi qu'aux processus qui permettent une participation plus équitable à la science. Cette approche contient selon moi les outils pour traiter le problème spécifique de la définition du trouble mental. Elle présente au moins deux avantages majeurs par rapport aux approches de la définition du trouble mental présentée plus tôt : 1) elle permet de rendre compte de l'interaction entre faits et valeurs dans la production de connaissances scientifiques et 2) elle permet de légitimer la présence de valeurs dans la production de connaissances scientifiques, si

ce processus est socialement objectif. Pour Longino, les valeurs sociales et culturelles présentes en science doivent être soumises à la critique de la communauté scientifique pour faire partie légitimement de la science. Ces éléments m'apparaissent centraux pour dépasser la dichotomie entre objectivisme des fonctions naturelles et normativisme quant à la définition du trouble mental. Le cadre théorique développé par Longino a le potentiel de sortir la philosophie de la psychiatrie de l'impasse dans laquelle elle se trouve au regard de la question de l'influence des valeurs sur le concept de trouble mental. Grâce à cette approche, je pourrai défendre qu'un concept de trouble mental socialement construit et objectif est possible, si les processus par lesquels le concept est défini respectent les standards de l'objectivité sociale.

Présentation générale de la thèse

Dans ce qui suit, je présenterai les différentes sections qui me permettront de défendre un concept socialement construit et objectif de trouble mental, fondé sur la théorie de l'objectivité sociale de Longino. Ma thèse sera découpée en sept chapitres. Les trois premiers chapitres consisteront en une exposition des différentes approches traditionnelles qui ont jusqu'à maintenant composé le débat philosophique sur la définition. Plus précisément, les deux premiers chapitres seront consacrés à un examen de la théorie biostatistique de la santé (BST), développée par Christopher Boorse. Le premier chapitre concernera la position de Boorse quant à la santé mentale et au trouble mental, alors que le deuxième chapitre visera la thèse de l'indépendance des valeurs des concepts de santé et de pathologie qu'il défend. Je me permets d'insister davantage sur les thèses de Boorse parce qu'elles ont été généralement moins bien comprises

lorsqu'appliquées en psychiatrie. Ces deux premiers chapitres me permettront donc dans un premier temps de présenter une interprétation plus juste de la BST lorsqu'elle concerne le trouble mental. Or j'adopterai également une posture critique vis-à-vis les thèses de Boorse. À la fin de ces deux chapitres, j'aurai montré que la BST n'est pas suffisante pour penser un concept objectif de trouble mental. Le troisième chapitre se veut une exploration des différentes alternatives à la BST. J'analyserai d'abord certaines approches normativistes, pour ensuite en montrer les limites. Je me pencherai ensuite sur une autre élaboration de l'approche objectiviste des fonctions naturelles, soit la théorie de la dysfonction préjudiciable (DP) de Jerome Wakefield. Bien que permettant de surmonter certains problèmes de la BST, la DP fait elle-même face à des obstacles de taille. Au terme de ces trois chapitres, une première conclusion partielle émergera : ces différentes approches présentent toutes des problèmes importants, et aucune ne permet, seule, de donner une définition pleinement adéquate du trouble mental.

Le chapitre 4 aura pour but de présenter de manière générale le cadre que je propose d'importer au sein du débat sur la définition du trouble mental. Dans un premier temps, je vais présenter une interprétation de l'approche objectiviste des fonctions naturelles empruntant à l'épistémologie sociale. L'approche objectiviste des fonctions naturelles se fonde sur l'idéal de la science indépendante des valeurs, soit la thèse selon laquelle une science objective (et donc légitime) est une science indépendante des valeurs. Je vais présenter différentes critiques ayant visé cet idéal pour ensuite proposer une alternative : la théorie de l'objectivité sociale. Le but de ce chapitre est de défendre que l'objectivité sociale apparaît comme un meilleur idéal pour la psychiatrie et son concept de trouble mental. Alors que ce chapitre adoptera une posture plutôt descriptive vis-à-vis la théorie de l'objectivité sociale, le chapitre 5 sera plus critique : je vais passer en revue différentes attaques ayant pris pour cible la théorie de l'objectivité sociale de Longino. Ce chapitre me permettra de solidifier mon argument en défendant

l'objectivité sociale face à certaines de ces critiques, mais aussi à proposer des amendements à la théorie de Longino dans les cas où les critiques paraissent légitimes. Je défendrai donc ultimement une version amendée de l'objectivité sociale capable de résister aux critiques l'ayant visée, ce que j'appellerai OS*. Ayant clarifié la notion d'objectivité sociale qui pourra permettre de penser la définition du trouble mental, dans le chapitre 6, j'examinerai la possibilité que le trouble mental soit à la fois socialement construit *et* empirique en revisitant la notion de dysfonction, caractéristique de l'approche objectiviste des fonctions naturelles, dans le cadre de l'OS*. Je défendrai l'idée selon laquelle le fait de reconnaître l'influence des valeurs sociales et culturelles sur la définition du trouble mental n'entraîne pas l'abandon de la notion de dysfonction, assurant ainsi le caractère empirique du trouble mental. Enfin, dans le chapitre 7, j'exporterai les conclusions des chapitres 4 et 5 en les appliquant aux structures de prise de décision de l'APA responsables de définir le trouble mental. Ce chapitre se veut une exploration de la manière d'opérationnaliser la théorie de l'objectivité sociale au cas concret de la définition du trouble mental. Je présenterai une ébauche de ce à quoi pourrait ressembler le processus qui mènerait à un concept de trouble mental socialement objectif au sein du DSM. Au terme de cette thèse, j'espère contribuer à la reconnaissance du potentiel de l'épistémologie sociale des sciences pour penser les enjeux de valeurs en psychiatrie, et plus spécifiquement à l'élaboration de processus de prise de décision quant à un concept scientifique et normatif comme le trouble mental.

Références de l'introduction

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders*. (3rd ed.). Washington : American Psychiatric Association.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders*. (3rd ed., revised). Washington : American Psychiatric Association.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. (4th ed.). Washington : American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. (4th ed., text revision). Washington : American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. (5th ed.). Washington : American Psychiatric Association.

Armstrong, T. (2010). *Neurodiversity : Discovering the extraordinary gifts of autism, ADHD, dyslexia, and other brain differences*. Cambridge, MA : Da Capo Press.

Aucouturier, V. et Demazeux, S. (2013). The concept of “mental disorder”. Dans H. Carel et R. Cooper (dir.), *Health, illness and disease* (p. 89-106). Londres : Routledge.

Bayer, R. (1981). *Homosexuality and American psychiatry : The politics of diagnosis*. New York : Basic Books.

Bayer, R. (1987). *Homosexuality and American psychiatry : The politics of diagnosis*. Princeton : Princeton University Press.

Bergner, R. M. (1997). What is psychopathology? And so what? *Clinical Psychology : Science and Practice*, 4(3), 235-248.

- Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, 3(2), 207-220.
- Bolton, D. (2001). Problems in the definition of “mental disorder”. *The Philosophical Quarterly*, 51(203), 182-199.
- Bolton, D. (2008). What is mental disorder? An essay in philosophy, science, and values. New York : Oxford University Press.
- Bueter, A. (2019). Social epistemology and psychiatry. Dans R. Bluhm et S. Tekin (dir.), *Bloomsbury companion to philosophy of psychiatry* (p.485-503). Londres : Bloomsbury Academic.
- Conrad, P. et Potter, D. (2000). From hyperactive children to ADHD adults : Observations on the expansion of medical categories. *Social Problems*, 47, 559-582.
- Conrad, P. (2007). *The medicalization of society*. Baltimore : Johns Hopkins University Press.
- Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(2), 263-282.
- Cooper, R. (2005). *Classifying madness : A philosophical examination of the diagnostic and statistical manual of disorders*. Dordrecht : Springer.
- Cooper, R. (2007). Can it be a good thing to be deaf? *The Journal of Medicine and Philosophy*, 32(6), 563-583.
- Cooper, R. (2015). Must disorders cause harm? The changing stance of the DSM. Dans S. Demazeux et P. Singy (dir.), *The DSM-5 in perspective* (p. 83-96). Dordrecht : Springer.

- Cooper, R. (2021) On harm. Dans L. Faucher et D. Forest (dir.), *Defining mental disorder : Jerome Wakefield and his critics* (p. 537-552). Cambridge, MA : MIT Press.
- Cuthbert, B. N. (2015). Research domain criteria : Toward future psychiatric nosologies. *Dialogues in Clinical Neuroscience*, 17(1), 89-97.
- Decker, H. S. (2007). How Kraepelinian was Kraepelin? How Kraepelinian are the neo-Kraepelinians? From Emil Kraepelin to *DSM-III*. *History of Psychiatry*, 18, 337-360.
- Decker, H. S. (2013). *The making of DSM-III : A diagnostic manual's conquest of American psychiatry*. Oxford : Oxford University Press.
- Demazeux, S. (2021). *DSM in the light of HDA (and conversely)*. Dans L. Faucher et D. Forest (dir.), *Defining mental disorders : Jerome Wakefield and his critics* (p. 3-26). Cambridge, MA : MIT Press.
- DeVito, S. (2000). On the value-neutrality of the concepts of health and disease : Unto the breach again. *The Journal of Medicine and Philosophy*, 25(5), 539-567.
- De Vreese, L. (2017). How to proceed in the disease concept debate? A pragmatic approach. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 42(4), 424-446.
- Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, 67(4), 559-579.
- Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, 138(3), 453-473.
- Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh : University of Pittsburgh Press.

- Dupré, J. (2007). Fact and value. Dans H. Kincaid, J. Dupré et A. Wylie (dir.), *Value-free science? Ideals and illusions* (p. 27-41). New York : Oxford University Press.
- Elliott, K. C. (2011). *Is a little pollution good for you? Incorporating societal values in environmental research*. New York : Oxford University Press.
- Engelhardt, H. T. (1975). The concepts of health and disease. Dans H. T. Engelhardt et S. F. Spicker (dir.), *Evaluation and explanation in the biomedical sciences* (p. 125-141). Springer, Dordrecht.
- Ereshefsky, M. (2009). Defining “health” and “disease”. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 40(3), 221-227.
- Faucher, L. et Forest, D. (2021). *Defining mental disorders : Jerome Wakefield and his critics*. Cambridge, MA : MIT Press.
- Fulford, K. W. M. (1989). *Moral theory and medical practice*. Cambridge : Cambridge University Press.
- Fulford, K. W. M. (1991). Evaluative delusions : Their significance for philosophy and psychiatry. *British Journal of Psychiatry*, 159, 108-112.
- Fulford, K. W. M. (2000). Teleology without tears : Naturalism, neo-naturalism, and evaluationism in the analysis of function statements in biology (and a bet on the twenty-first century). *Philosophy, Psychiatry, & Psychology*, 7(1), 77-94.
- Fulford, K. W. M. (2001). “What is (mental) disease?” : An open letter to Christopher Boorse. *Journal of Medical Ethics*, 27(2), 80-85.
- Fulford, K. W. M. (2005). Values in psychiatric diagnosis : Developments in policy, training and research. *Psychopathology*, 38, 171-176.
- Fuller, S. (1988). *Social epistemology*. Bloomington, IN : Indiana University Press.

- Gamwell, L. et Tomes, N. (1995). *Madness in America : Cultural and medical perspectives of mental illness before 1914*. Ithaca, NY : Cornell University Press.
- Garson, J. R. (2006). *Psychiatric disorders and biological dysfunctions : Some philosophical questions concerning psychiatry*. (Thèse de doctorat non publiée). University of Texas at Austin.
- Goffman, E. (1961). *Asylums : Essays on the social situations of mental patients and other inmates*. Oxford : Doubleday.
- Goldman, A. et Blanchard, T. (2018). Social Epistemology. Dans E. N. Zalta (dir.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/fall2019/entries/epistemology-social/>
- Grob, G. N. (1994b). *The mad among us : A history of the care of America's mentally ill*. Cambridge, MA : Harvard University Press.
- Grob, G. N. (2011). The attack of psychiatric legitimacy in the 1960s : Rhetoric and reality. *Journal of the History of the Behavioral Sciences*, 47(4), 398-416.
- Heinrichs, R. W. (2001). *In search of madness : Schizophrenia and neuroscience*. New York : Oxford University Press.
- Horwitz, A. V. et Wakefield, J. C. (2007). *The loss of sadness : How psychiatry transformed normal sorrow into depressive disorder*. New York : Oxford University Press.
- Hyman, S. E. (2010). The diagnosis of mental disorders : The problem of reification. *Annual Review of Clinical Psychology*, 6, 155-179.
- Jablensky, A. (2007). Living in a Kraepelinian world : Kraepelin's impact on modern psychiatry. *History of Psychiatry*, 18, 381-388.
- Kincaid, H. (2007). Contextualist morals and science. Dans H. Kincaid, J. Dupré et A. Wylie (dir.), *Value-free science : Ideals and illusions?* (p. 218-238). Oxford : Oxford University Press.

- Kingma, E. (2013). Naturalist accounts of mental disorder. Dans K. W. M. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini et T. Thornton (dir.), *The Oxford handbook of philosophy and psychiatry* (p. 363-384). Oxford : Oxford University Press.
- Kirk, S. A. et Kutchins, H. (1992). *The selling of DSM : The rhetoric of science in psychiatry*. New York : Routledge.
- Kutchins, H. et Kirk, S. A. (1997). *DSM : The psychiatric bible and the creation of mental disorders*. New York : The Free Press.
- Lane, C. (2007). *Shyness : How normal behavior became a sickness*. New Haven, CT : Yale University Press.
- Lemoine, M. (2021) Is the dysfunction component of the “harmful dysfunction analysis” stipulative? Dans L. Faucher et D. Forest, *Defining mental disorder : Jerome Wakefield and his critics* (p. 199-212). Cambridge, MA : MIT Press.
- Longino, H. E. (1990). *Science as social knowledge : Values and objectivity in scientific inquiry*. Princeton, NJ : Princeton University Press.
- Longino, H. E. (2002). *The fate of knowledge*. Princeton, NJ : Princeton University Press.
- Mayes, R. et Horwitz, A. V. (2005). DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences*, 41, 249-267.
- Murphy, D. (2006). *Psychiatry in the scientific image*. Cambridge, MA : MIT Press.
- Organisation mondiale de la Santé (1946). Préambule à la constitution de l'Organisation mondiale de la Santé, tel qu'adopté par la Conférence internationale sur la Santé, New York, 19 juin-22 juillet 1946; signé le 22 juillet 1946 par les représentants de 61 États. Dans *Actes officiels de l'Organisation mondiale de la Santé*, 2, 100. Genève : OMS.

- Potter, N. N. (2012). Mad, bad, or virtuous? The moral, cultural, and pathologizing features of defiance. *Theory & Psychology*, 22(1), 23-45.
- Potter, N. N. (2014). Oppositional defiant disorder : Cultural factors that influence interpretations of defiant behavior and their social and scientific consequences. Dans H. Kincaid et J. A. Sullivan (dir.), *Classifying psychopathology : Mental kinds and natural kinds* (p. 175). Cambridge, MA : MIT Press.
- Porter, R. (2002). *Madness : A brief history*. Oxford : Oxford University Press.
- Putnam, H. (2002). *The collapse of the fact/value dichotomy*. Cambridge, MA : Harvard University Press.
- Radden, J. (2019). Mental Disorder (Illness). Dans E. N. Zalta (dir.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/win2019/entries/mental-disorder/>
- Ross, P. A. (2002). The fact value dichotomy in demarcating disorder. *Philosophy, Psychiatry, & Psychology*, 14(2), 107-109.
- Sadegh-Zadeh, K. (2008). The prototype resemblance theory of disease. *Journal of Medicine and Philosophy*, 33(2), 106-139.
- Schmitt, F. F. (dir.). (1994). *Socializing epistemology : The social dimensions of knowledge*. Lanham, MA : Rowman & Littlefield.
- Schuur, R. (2019). Mental health and illness : Past debates and future directions. Dans S. Tekin et R. Bluhm (dir.), *The Bloomsbury Companion to Philosophy of Psychiatry* (p. 527-541). Londres : Bloomsbury Academic.
- Schwartz, Peter H. (2007). Defining dysfunction : Natural selection, design, and drawing a line. *Philosophy of Science*, 74(3), 364-385.
- Shields, K. et Beversdorf, D. (2020). A dilemma for neurodiversity. *Neuroethics*.
<https://doi.org/10.1007/s12152-020-09431-x>

- Shorter, E. (1997). *A history of psychiatry*. New York : John Wiley and Sons.
- Shorter, E. (2013). The history of the DSM. Dans J. Paris et J. Phillips (dir.), *Making the DSM-5 : Concepts and controversies* (p. 3-19). New York : Springer.
- Silberman, S. (2015). *NeuroTribes : The legacy of autism and the future of neurodiversity*. New York : Avery.
- Spitzer, R. L. et Endicott, J. (1978). Medical and mental disorder : Proposed definition and criteria. Dans R. L. Spitzer et D. F. Klein (dir.). *Critical issues in psychiatric diagnosis* (p. 15-39). New York : Raven Press.
- Szasz, T. (1961), *The myth of mental illness : Foundations of a theory of personal conduct*. New York : Harper & Row.
- Tabb, K. (2015). Psychiatric progress and the assumption of diagnostic discrimination. *Philosophy of Science*, 82, 1047-1058.
- Tekin, Ş. (2019). The missing self in scientific psychiatry. *Synthese*, 196, 2197-2215.
- Thornton, T. (2007). *Essential philosophy of psychiatry*. Oxford : Oxford University Press.
- Tsou, J. Y. (2011). The importance of history for philosophy of psychiatry : The case of the DSM and psychiatric classification. *Journal of the Philosophy of History*, 5, 446-470.
- Tsou, J. Y. (2015). DSM-5 and psychiatry's second revolution : Descriptive vs. theoretical approaches to psychiatric classification. Dans S. Demazeux e P. Singy (dir.). *The DSM-5 in perspective* (p. 43-62). Dordrecht : Springer.
- Tyrer, P. et Steinberg, D. (2006). *Models for mental disorder : Conceptual models in psychiatry*. (4th edition). West Sussex : John Wiley & Sons.
- Varga, S. (2011). Defining mental disorder : Exploring the "natural function" approach. *Philosophy, Ethics, and Humanities in Medicine*, 6(1), 1-10.

- Varga, S. (2017). Mental disorder between naturalism and normativism. *Philosophy Compass*, 12(6), 1-9.
- Wakefield, J. C. et First, M. B. (2003). Clarifying the distinction between disorder and non-disorder : Confronting the overdiagnosis (“false positives”) problem in DSM-V. Dans K. A. Phillips, M. B. First et H. A. Pincus (dir.), *Advancing DSM : Dilemmas in psychiatric diagnosis* (p. 23-56). Arlington, VA : American Psychiatric Publisher.
- Wilson, M. (1993). DSM-III and the transformation of American psychiatry : A history ». *American Journal of Psychiatry*, 150, 399-410.

CHAPITRE 1

LA THÉORIE BIOSTATISTIQUE DE CHRISTOPHER BOORSE

Introduction

Christopher Boorse est l'un des philosophes anglo-saxons de la médecine les plus connus. Sa notoriété provient des thèses audacieuses qu'il a défendues dans le débat entre naturalisme et normativisme quant à la définition de la santé et de la maladie. L'ambition de Boorse, qui se révèle dans une série d'articles publiés à partir de 1975, consiste à produire une définition de la santé et de la maladie qui soit indépendante des valeurs (*value-free*). L'analyse de Boorse des concepts théoriques de santé et de pathologie se présente sous la forme de ce qu'il appelle la théorie biostatistique de la santé (*biostatistical theory of health*, l'abréviation est BST). Cette théorie prend ancrage dans un langage fonctionnel qui permettrait, selon Boorse, de fournir les bases scientifiques et empiriques requises aux concepts de santé et de pathologie en médecine. Dans ce cadre, la santé est associée au fonctionnement normal, tandis que la pathologie l'est à un dysfonctionnement interne présent chez un individu. Dans la section 1.1, je vais présenter les grandes lignes de la BST. La BST est une théorie complexe, qui a été largement critiquée et commentée (voir l'introduction de la thèse). Dans cette section, je vais m'intéresser à trois aspects centraux de la BST : l'idée de fonction normale (section 1.1.1), les notions de normalité statistique et de classe de

référence (section 1.1.2) et la méthodologique utilisée par Boorse pour parvenir à la BST (section 1.1.3). Après avoir clarifié ces différents éléments, je vais m'intéresser à un aspect de la BST ayant reçu moins d'intérêt dans les écrits : son application à la psychiatrie. Bien que la psychiatrie ait été le premier intérêt de Boorse, le seul article portant exclusivement sur ce sujet est *What a theory of mental health should be?* (1976). Dans ses publications ultérieures, le thème de la santé mentale n'a reçu que très peu d'attention de la part de Boorse, parce que c'est dans le domaine de la médecine somatique que les critiques de la BST ont été les plus nombreuses (Boorse, 1997). Or, c'est lorsque l'on s'est intéressé à l'application de la BST à la psychiatrie que la position de Boorse a été généralement moins bien comprise. Étant donné que le sujet de cette thèse est le concept de trouble mental, je vais donc m'attarder à reconstruire la manière dont Boorse conçoit le concept de trouble mental dans le cadre de la BST dans la section 1.2. Cette analyse permettra de révéler que Boorse est beaucoup plus proche des thèses du mouvement antipsychiatrique que ce qu'on a laissé entendre, principalement quant à l'absence d'un concept théorique de santé mentale au sein des théories psychiatriques des années 1970. Après avoir mis en relief cette thèse de Boorse (sections 1.2.1 à 1.2.5), dans la section 1.2.6, je vais appliquer la BST à la classification officielle des troubles mentaux de la psychiatrie américaine, le DSM-5. Cette application vise à déterminer si Boorse défendrait l'existence d'un concept théorique de trouble mental aujourd'hui. Mon analyse révélera que le DSM-5 ne présente pas de concepts théoriques de santé mentale, au sens entendu par Boorse. Ce chapitre vise donc à mieux comprendre la position de Boorse quant au concept spécifique de santé mentale.

1.1 Présentation générale de la théorie biostatistique de Boorse

Pour bien comprendre le projet de Boorse, il faut distinguer deux significations qui peuvent être intuitivement associées au concept général de santé : une signification pratique et une signification théorique (Boorse, 1975). Selon Boorse, le concept pratique de santé renvoie généralement au caractère subjectif souvent indésirable de la pathologie, à l'aspect thérapeutique ainsi qu'aux normes sociales et culturelles qui influent sur la signification du concept. Cette signification est donc liée à l'aspect clinique et évaluatif de la médecine. Avant la contribution de Boorse, c'était souvent la seule signification du concept de santé qui a été retenue. Or selon Boorse, le concept de santé possède également une signification théorique distincte de la signification pratique, signification qui n'avait jusqu'alors pas été analysée en philosophie de la médecine anglo-saxonne. Boorse distingue donc deux projets associés à ces deux significations distinctes : celui de rendre compte de la signification de la maladie (*illness*) dans le cadre de la pratique médicale et celui d'analyser le concept de pathologie (*pathological condition*) au sein de la théorie médicale. C'est ce second projet qui occupe Boorse. Celui-ci propose ainsi d'analyser les concepts de santé et de pathologie *tels qu'ils sont compris par la physiologie*, puisque c'est cette discipline qui assurerait à la médecine clinique un fondement théorique, et donc qui pourrait fournir un concept de pathologie indépendant des valeurs.

Cette clarification faite, il convient de dresser un portrait général de la BST, dont les grands traits ont été rassemblés par Boorse sous la forme d'un résumé en 1997 :

1. The reference class is a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species.

2. A normal function of a part or process within members of the reference class is a statistically typical contribution by it to their individual survival [or] reproduction.
3. Health in a member of the reference class is normal functional ability: the readiness of each internal part to perform all its normal functions on typical occasions with at least typical efficiency.
4. A disease [later, pathological condition] is a type of internal state which impairs health, i.e., reduces one or more functional abilities below typical efficiency (Boorse, 1997, p. 7-8).

Le concept théorique de santé serait donc la conformité des états internes de l'organisme au *design* fonctionnel typique de sa classe de référence. La pathologie serait définie comme une déviation par rapport au *design* fonctionnel typique de la classe de référence à laquelle appartient l'organisme. Dans ce qui suit, je vais approfondir cette caractérisation en m'attardant aux notions de fonction normale, de normalité statistique et de classe de référence sur lesquelles se fonde la BST. Je vais ensuite m'intéresser à la méthode par laquelle Boorse parvient à cette définition de la santé et de la pathologie. Tous ces éléments me semblent pertinents pour bien comprendre les nuances de la position de Boorse quant au concept théorique de trouble mental en psychiatrie ainsi que ses thèses sur la possibilité que ce concept soit indépendant des valeurs.

1.1.1 Le concept de fonction normale chez Boorse

La notion de fonction est l'un des constituants centraux de la BST. Cette notion permettrait à Boorse d'affirmer que sa théorie est indépendante des valeurs, étant donné qu'il s'agit d'une notion biologique, ou comme Boorse le dit, d'une norme naturelle (Boorse, 1977, p. 554). Boorse précise sa conception de la fonction comme une

contribution à un but, s'inscrivant dans l'approche des fonctions comme « orientées vers un but » (*goal-directed function*) (Boorse, 1976b, 1977, 2002; voir aussi Nagel, 1961; Sommerhoff, 1950 pour la même approche des fonctions). Selon Boorse, la physiologie décrit les systèmes biologiques comme des systèmes téléologiques. L'organisme est alors représenté comme un système total divisé en sous-systèmes, partant des cellules aux organes, jusqu'au comportement total de l'organisme. Boorse défend donc un « fonctionnalisme des parties » (*part-functionalism*, Boorse, 1997), au sens où l'organisme se voit décomposé en différents modules selon des niveaux de complexité biologiques hiérarchiquement différents, chacun de ces modules possédant une fonction propre relativement au tout de l'organisme. Dans cette perspective, la fonction d'une partie ou d'un processus serait sa contribution causale aux buts supérieurs de l'organisme. En physiologie, les buts de l'organisme seraient la survie et la reproduction des organismes individuels. Les processus et les parties de ces sous-systèmes possèdent alors des fonctions « naturelles » s'ils contribuent de manière causale aux buts hiérarchiquement supérieurs du système que sont la survie et la reproduction.

Par exemple, un organe comme le cœur aurait pour fonction de faire circuler le sang étant donné que la circulation du sang contribue causalement à la survie de l'organisme. La circulation du sang serait donc la fonction biologique du cœur, par contraste avec le fait, pour le cœur, de produire un battement sonore lorsqu'il fait circuler le sang, qui ne serait qu'un simple effet corrélé de la fonction. Selon Boorse, il est en effet possible d'affirmer que c'est la fonction biologique du cœur de pomper le sang, étant donné que pomper le sang contribue causalement à la survie de l'organisme, ce qui n'est pas le cas du battement sonore du cœur (Boorse, 1977, p. 556, l'argument provient initialement de Nagel, 1961 et est repris par Boorse).

1.1.2 Normalité statistique et classe de référence chez Boorse

Selon Boorse, cette conception de la fonction ne peut constituer à elle seule le concept théorique de santé. Supposons que le bruit produit par le battement du cœur contribue, chez un organisme individuel, à sa survie et à sa reproduction (par exemple, en alarmant un docteur si celui-ci arrête de produire le battement sonore, permettant ainsi au docteur de sauver l'individu de la mort). Il n'en découlerait pas pour autant que ce soit la fonction biologique du cœur que de produire un battement sonore. En guise d'illustration, Boorse donne l'exemple d'un écureuil dont la queue, prise dans un piège, lui permettrait d'éviter de se faire frapper par une voiture sur l'autoroute, contribuant ainsi causalement à la survie et à la reproduction de l'organisme. Cela n'impliquerait pas pour autant que la queue de l'écureuil a comme fonction biologique la défense contre les voitures (Boorse, 1977, p. 557). Il s'en suit une nécessité de préciser la notion de fonction biologique. Selon Boorse, l'attribution de fonction en physiologie ne concerne pas les individus particuliers, mais les populations, ou l'espèce. Cela s'explique par le fait que la physiologie est intéressée davantage par des fonctions *typiques de l'espèce*, et non par des fonctions individuelles. Ainsi, en physiologie, l'attribution de fonction en termes de contribution causale à la survie et à la reproduction fait référence à la contribution *standard* des parties de l'organisme relativement à la *classe de référence* à laquelle appartient l'organisme. Sans ces précisions, il ne serait pas possible de distinguer les fonctions physiologiques normales des effets accidentels des parties de l'organisme qui contribuent à la survie et à la reproduction.

Ainsi, pour Boorse, une fonction biologique normale doit être comprise comme une contribution causale typique des parties ou des processus de l'organisme vers les buts

hiérarchiquement supérieurs de l'organisme, la survie ou la reproduction. La typicité des fonctions biologiques est établie par idéalisation statistique :

But the field naturalist abstracts from individual differences and from disease by averaging over a sufficiently large sample of the population. The species design that emerges is an empirical ideal which, I suggest, serves as the basis for health judgments in any species where we make such judgments. (Boorse, 1977, p. 557).

Les fonctions biologiques typiques sont donc déterminées par mesure statistique, faisant de la norme statistique une notion empirique. Ainsi, défendre les écureuils contre les voitures n'est pas la fonction biologique de la queue de l'écureuil, car la queue de l'écureuil ne contribue pas *de manière statistiquement typique* à la survie et à la reproduction des écureuils. Selon Boorse, pour qu'une analyse du concept théorique de santé soit complète, la notion de fonction comme orientée vers un but ne saurait donc se passer de cette idée de typicité. Boorse introduit alors la notion de *design fonctionnel* pour référer à cette typicité fonctionnelle (Boorse, 1977, p. 555). Le fonctionnement des parties d'un organisme individuel ne peut être dit normal que s'il correspond au *design* fonctionnel de l'espèce, pointant vers les fonctions *statistiquement* typiques au sein de l'espèce. En ce sens, un membre de l'espèce correspond à ce qui est fonctionnellement typique si chacune de ses parties contribue de manière statistiquement typique à la survie et à la reproduction de l'organisme, selon un niveau d'efficacité statistiquement typique. Le *design* fonctionnel, qui correspond à la norme biologique chez Boorse, est donc établi selon une norme statistique.

Boorse ajoute une dernière notion qui viendrait compléter la signification de la santé et de la pathologie en physiologie. Les physiologistes ne font pas seulement référence à l'espèce lorsqu'ils font usage de l'idée de *design* fonctionnel, mais bien à des sous-groupes d'individu de *design* fonctionnel uniforme au sein de l'espèce. En guise d'exemple, Boorse parle des différences fonctionnelles entre les systèmes

reproducteurs des hommes et des femmes et des différentes fonctions associées aux étapes de développement de l'enfance. Dans ces deux cas, il semble y avoir des différences fonctionnelles importantes entre des sous-groupes de l'espèce humaine. Cette hétérogénéité fonctionnelle légitimerait le découpage de la classe de référence qu'est l'espèce en sous-classes de référence. Le sexe, l'âge, voire la race, constitueraient les classes de référence utilisées par la physiologie pour rendre compte de ce type d'hétérogénéité fonctionnelle au sein de l'espèce.

Pour résumer, selon la BST, un organisme peut être dit en santé si l'ensemble de ses parties ou processus accomplissent leur fonction typique, avec un niveau d'efficacité statistiquement typique, conformément au *design* fonctionnel de la classe de référence à laquelle appartient l'organisme. La pathologie est en ce sens une altération du fonctionnement normal par rapport à la norme donnée par la classe de référence, c'est-à-dire une réduction d'une ou plusieurs des capacités fonctionnelles typiques en dessous du niveau d'efficacité typique. Pour Boorse, ces concepts théoriques de pathologie et de santé sont indépendants des valeurs.

1.1.3 Le naturalisme méthodologique de Boorse

Après avoir présenté les grandes lignes de la BST, il convient de s'attarder à la méthode utilisée par Boorse pour parvenir à la définition des concepts théoriques de santé et de pathologie et à la manière dont Boorse peut être dit « naturaliste ». L'examen de la méthode utilisée par Boorse aura des répercussions sur l'interprétation de la position de Boorse quant au concept théorique de santé mentale et sur l'idée d'une définition indépendante des valeurs (sur laquelle je me penche davantage au chapitre 2). De

manière générale, en philosophie, le terme « naturalisme » réfère au courant de pensée qui pose que la réalité ne contient rien de supranaturel et que la science est la meilleure méthode pour appréhender cette réalité (Papineau, 2009). Il convient cependant de noter que la signification du naturalisme en philosophie de la médecine n'est pas aussi claire et est un point de débat (Giroux, 2016). En effet, ce qui a opposé les naturalistes et les normativistes dans le débat concernant la définition de la santé a porté davantage sur la présence de composantes normatives ou factuelles au sein de la définition : les normativistes ont défendu l'idée que la santé se définit en recourant à des valeurs et des normes sociales et culturelles, tandis que les naturalistes ont défendu que la santé est définie sans recours à des notions normatives. Boorse lui-même a adopté le terme naturalisme, mais en admettant qu'il aurait préféré que sa position soit nommée « descriptivisme » ou « non-normativisme » (Boorse, 1997, p. 4). C'est donc cette forme spécifique de naturalisme telle qu'elle s'est présentée en philosophie de la médecine et de la psychiatrie qui va m'intéresser dans cette section, de même que la méthodologie qu'elle présuppose chez Boorse.

Une attention grandissante a été portée à la méthodologie employée par les auteurs quant à la définition théorique des concepts de santé et de maladie (De Vreese, 2017; Lemoine, 2013; Lemoine et Giroux, 2016; Lilienfeld et Marino, 1995; Murphy, 2005; Murphy et Woolfolk, 2000a; Nordby, 2006; Sadegh-Zadeh, 2000, 2008; Schwartz, 2004, 2007, 2014). Ce qui ressort de ces écrits est la thèse selon laquelle la majorité des auteurs qui ont pris part au débat quant à la définition de la santé et de la maladie ont fait usage de *l'analyse conceptuelle* pour tenter de définir le concept. Boorse fait partie de ces auteurs, mais adopte une variante de l'analyse conceptuelle, la *reconstruction rationnelle*. Pour comprendre le lien entre le naturalisme de Boorse et sa reconstruction rationnelle, je vais me baser sur le travail Lemoine et Giroux (2016). Comme Lemoine et Giroux l'ont montré, Boorse adopte deux formes de naturalisme : un *naturalisme méthodologique* et un *naturalisme ontologique*. Le naturalisme méthodologique se caractérise par deux thèses principales : 1) la primauté accordée aux

concepts théoriques, scientifiques et indépendants des valeurs de la santé et de la pathologie et 2) l'utilisation de la méthode de l'analyse conceptuelle pour définir les concepts théoriques de santé et de maladie (Lemoine et Giroux, 2016, p. 24). Pour sa part, en philosophie de la médecine, le naturalisme ontologique comprend deux éléments centraux : 1) le naturisme et 2) la distinction naturelle. Le premier élément réfère à l'idée selon laquelle la santé est « une tendance naturelle » des organismes; tandis que le second renvoie à la thèse selon laquelle la distinction entre la santé et la pathologie serait un fait naturel. Ainsi, le naturalisme ontologique concerne le contenu de la réalité, et plus spécifiquement dans le contexte de la philosophie de la médecine, la croyance selon laquelle la santé est un phénomène qui existe dans la nature, indépendamment des valeurs et intérêts humains. Ici, c'est moins le contraste entre ce qui serait naturel et ce qui serait supranaturel qui importe que le contraste entre ce qui est un fait naturel et ce qui est de l'ordre des valeurs sociales ou culturelles.

Traditionnellement, la méthode de l'analyse conceptuelle cherche à décrire la signification d'un terme selon les intuitions des individus quant à ce terme ou selon la manière dont le terme est utilisé par les locuteurs. Cette méthode a généralement été employée pour définir des concepts clés en philosophie, tels que « connaissance », « justice », « liberté », etc. L'analyse conceptuelle cherche à mettre en lumière la signification des concepts selon des critères définitionnels nécessaires et suffisants. Par exemple, pour le cas de la connaissance, la définition classique est une « croyance vraie justifiée ». Par sa nature descriptive, la méthode de l'analyse conceptuelle exclut la stipulation de son analyse, c'est-à-dire que l'objectif est de révéler ce que la signification d'un terme est, et non pas ce qu'elle devrait être. Or dans le cadre du naturalisme méthodologique de Boorse, une première différence à noter avec la méthode classique de l'analyse conceptuelle est qu'à certaines occasions, Boorse est prêt à s'écarter de l'usage scientifique des termes de santé et de pathologie pour proposer un « meilleur » usage. Par exemple, Boorse discute du fait que parfois,

certain pathologistes considéreraient la fièvre et l'inflammation comme des pathologies. Boorse n'hésite pas à qualifier cet usage « d'erreur » :

A good example of what I mean by “considered” judgments is pathology’s view of defense mechanisms like inflammation, fever, and the immune response. Nordenfelt suggests that these are generally seen as part of a disease (Nordenfelt, 1987, 30-31; Nordenfelt, 2001, 17). I agree that inflammation and fever are often called pathological in medical books. However, I think this is an error, due to two causes. One is a traditional failure to recognize the nature of at least the first two of these processes as defense mechanisms, *i.e.*, normal, usually beneficial biological functions. For fever, the importance of this mistake has been emphasized by Wakefield (1999a, 392-393) and apostles of Darwinian medicine. The second cause is terminological: the lack of a standard medical term for “indicating disease.” I have proposed “pathodictic” for this purpose (Boorse, 1987, 387). If “pathodictic” joined “pathologic” and “pathogenic” in medical usage, I feel sure that pathologists would agree with me that in infection, inflammation, fever, and the immune response are pathodictic but normal (nonpathologic). (Boorse, 2014, p. 32)

Il est donc plus juste, comme le note Boorse lui-même (2011, p. 20), de dire que sa méthode est celle de la *reconstruction rationnelle* des concepts scientifiques, une variante de l'analyse conceptuelle où des stipulations mineures sont permises dans le but de donner une définition cohérente des concepts. Les outils de classification des pathologistes ne sont donc pas la seule autorité en ce qui a trait à la bonne extension du concept; la cohérence en est un autre. Pour Boorse, il convient plutôt de *reconstruire* le jugement réfléchi des pathologistes quant à la santé et la pathologie, à partir des conditions pathologiques classifiées dans les manuels diagnostiques. Cette reconstruction peut impliquer un léger révisionnisme vis-à-vis certaines catégories diagnostiques, tant que cela ne touche pas des conditions saines ou pathologiques paradigmatiques. Par exemple, une définition de la santé qui n'accommoderait pas le cancer par ses critères définitionnels ne serait pas une bonne analyse du concept. À

l'inverse, une définition qui ferait de la grossesse une maladie ne serait pas non plus une bonne analyse du concept.

Par ailleurs, en lien avec la première thèse du naturalisme méthodologique, pour parvenir à une définition théorique du terme, Boorse pose que c'est l'usage du concept par les experts qui doit être analysé, et non pas l'usage ordinaire, ce qui distingue également son approche de la méthode classique de l'analyse conceptuelle. En effet, en général, la méthode de l'analyse conceptuelle fait appel aux intuitions de tous pour déterminer la signification d'un concept. La méthode préconisée par Boorse fait appel aux critères d'utilisation du terme par les experts qui, pour le cas de la santé et la pathologie, sont les physiologistes. Les outils de référence sont alors les manuels de physiologie. Parce que c'est l'utilisation du terme par les physiologistes qui fixe la signification des concepts de santé et de maladie, Boorse pense que l'utilisation de cas hypothétiques pour tester la validité de la théorie a moins de valeur que les cas réels de pathologie :

I see no interesting or authoritative ordinary usage of "disease" that we need consult in analyzing it. A corollary is that hypothetical cases have less force on this topic than is usual in philosophy. Normally, hypothetical cases play a huge role in philosophical analysis. But if we seek to analyze a technical concept of medicine, we should not appeal to what "we" would call a disease, or pathological, unless "we" can predict hypothetical medical usage. (Boorse, 1997, p. 62)

Ainsi, pour déterminer si la BST est une bonne théorie de la santé et de la pathologie, le critère est la correspondance de la théorie avec les outils de classification de la physiologie, et non pas l'utilisation d'expériences de pensée ou les intuitions des non-experts (voir aussi Boorse, 1977, p. 563).

Ainsi, par la méthode de la reconstruction rationnelle, Boorse prétend offrir une définition théorique de la santé et de la pathologie qui reconstruit le mieux l'usage que les pathologistes en font dans leur jugement, usage qui se révèle dans les manuels de classification. Bien que la pertinence et la portée de la méthode de l'analyse conceptuelle soient de plus en plus remises en question, au regard de l'objectif de ce chapitre, je vais laisser de côté ces critiques pour me concentrer sur l'impact d'une telle méthode sur la conception de la santé mentale que Boorse défend.

1.2 L'application de la BST à la psychiatrie

Dans la section précédente, j'ai mis en lumière les éléments clés qui permettent de comprendre la BST et la méthodologie employée par Boorse pour parvenir à une telle théorie. Dans cette section, je me pencherai sur la manière dont le concept spécifique de santé mentale s'intègre à la BST. Qui plus est, je vais m'intéresser à l'opposition que l'on trace généralement entre Boorse et les auteurs du mouvement antipsychiatrique. Dans ses articles sur la santé mentale, Boorse identifie Thomas Szasz et Theodore Sarbin, deux auteurs du mouvement antipsychiatrique, comme les représentants exemplaires de l'approche normativiste. Comme mentionné en introduction, Boorse est par opposition considéré comme le défenseur par excellence de l'approche naturaliste de la définition de la santé et de la pathologie (p. ex. : Kingma, 2013). Il est vrai que la démarche de Boorse, dans la série d'articles qu'il publie à partir de 1975, s'inscrit dans l'approche naturaliste comme je l'ai mise en lumière dans la section précédente, et qu'il se positionne bel et bien contre certaines des attaques normativistes ayant visé la psychiatrie. Dans cette section, je vais cependant nuancer cette classification en montrant en quel sens Boorse se dit naturaliste/objectiviste vis-

à-vis les concepts centraux de la psychiatrie. La conception de Boorse du concept de trouble mental est beaucoup plus critique de l'entreprise psychiatrique que ce qui a été mis de l'avant, et les similarités entre Boorse d'un côté, et Szasz et Sarbin de l'autre, beaucoup plus importantes qu'on pourrait le croire à première vue. Qui plus est, la grande majorité des commentateurs de Boorse présentent sa position quant au concept de santé mentale comme une simple extension de la BST. Selon cette interprétation, la conception de Boorse de la santé mentale ne serait qu'un simple calque de sa conception de la santé physique. Or cette vue est erronée : l'analyse du concept de santé mentale que Boorse offre diffère en plusieurs points de son analyse du concept de santé somatique. Malgré le fait que la position de Boorse soit complexe et touche à plusieurs enjeux entourant le concept de santé mentale, il m'apparaît crucial de m'arrêter sur deux des thèses principales de Boorse étant donné qu'elles auront un impact sur la classification de Boorse comme un naturaliste. La première thèse concerne l'absence d'un concept théorique de santé mentale dans les théories psychiatriques de l'époque, et la seconde touche à l'adoption du paradigme biomédical et physiologique en psychiatrie. La première thèse place Boorse beaucoup plus proche de Szasz et Sarbin, tandis que la seconde constitue l'essentiel du désaccord qui oppose Boorse aux deux auteurs.

Pour bien faire apparaître cette nuance, il convient de distinguer deux questions posées par Boorse : « est-il possible de parvenir à un concept théorique de santé mentale qui satisferait la BST? » et « est-ce la psychiatrie possède un tel concept? ». Boorse répond par l'affirmative à la première question, mais par la négative à la seconde. Boorse est donc bel et bien naturaliste quant à la *possibilité* de parvenir à un concept théorique de santé mentale analogue à celui retrouvé en physiologie, mais très près de certains antipsychiatres quant à l'*existence* d'un tel concept au sein de la psychiatrie de son époque. En effet, dans ses articles des années 1970, Boorse donne en grande partie raison aux critiques de l'antipsychiatrie. Selon Boorse, si l'on analyse le ou les concepts de santé mentale disponibles, il faudrait convenir qu'ils ne reposent pas sur

le *design* fonctionnel mental de l'espèce, et qu'en ce sens, ils ne sont pas conformes aux exigences d'un concept théorique de santé tel qu'il le propose avec la BST (Boorse, 1976). Ce que Boorse défend est plutôt qu'il n'est pas impossible, *en principe*, de parvenir à la connaissance du fonctionnement normal de l'esprit humain.

Le second argumentaire de Boorse qui m'intéressera dans cette section est celui de la recommandation de l'élaboration d'un concept théorique de santé mentale conforme à la BST comme trouvée en physiologie. Selon Boorse, la psychiatrie serait confrontée à un choix : elle doit soit abandonner le cadre biomédical et changer de vocabulaire, renonçant à l'utilisation des concepts de santé et de maladie, soit développer une théorie du fonctionnement mental typique de l'humain, ce qui lui permettrait de légitimer son utilisation du vocabulaire médical. Boorse penche vers cette seconde option, selon laquelle la psychiatrie devrait viser à développer une théorie de la santé mentale suivant un modèle fonctionnel. C'est donc en ce sens que Boorse s'écarte des thèses de l'antipsychiatrie, en ce qu'il croit *possible et souhaitable* l'élaboration d'une telle théorie, conformément à la BST en physiologie. Les antipsychiatres, quant à eux, prescrivent le retrait de la psychiatrie de l'entreprise médicale et, avec ce retrait, l'abandon du vocabulaire propre à la médecine. Si l'on qualifie Boorse de naturaliste quant au concept de trouble mental, ce doit donc être relativement à la thèse de la possibilité et de la désirabilité d'élaborer un concept théorique de santé mentale, et non de la présence d'un concept théorique de santé mentale au sein des théories psychiatriques de l'époque. C'est ce point qui le distingue clairement de l'antipsychiatrie, et qui distingue la BST lorsqu'appliquée au concept de santé somatique et au concept de santé mentale.

1.2.1 Le contexte de la publication de l'article

Pour bien comprendre la position de Boorse quant au concept de santé mentale, il faut chercher dans un article précis, *What a Theory of Mental Health Should be?*, publié en 1976. La majeure partie des thèses de Boorse quant à la santé mentale s'y trouve. Il peut être utile de situer l'article dans son contexte, car il s'agit d'un article qui vise des cibles précises, principalement certains auteurs du mouvement antipsychiatrique comme Szasz et Sarbin, ainsi que les théoriciens béhavioristes qui occupent une place dominante dans la psychologie de l'époque. En effet, Boorse s'adresse aux antipsychiatres que sont Szasz et Sarbin en ce qu'ils soutiennent que le concept de maladie mentale ne serait qu'un mythe, voire le simple reflet des normes sociales, culturelles et politiques (voir l'introduction pour une présentation plus générale du mouvement). Boorse vise également les théoriciens béhavioristes qui acceptent explicitement que le concept de maladie mentale désigne un trouble du comportement, où un trouble mental ne serait que le reflet d'une inadaptation de l'individu face à la société. Ces deux approches auraient en commun le fait qu'elle nie l'existence de bases théoriques au concept de santé mentale, bases théoriques qui viendraient assurer la légitimité de la psychiatrie en ce qu'elle aurait de scientifique (Boorse, 1976, p. 61). Dans cette mise en contexte, je vais brièvement présenter les idées de Szasz et Sarbin, puisque l'on a souvent opposé ces auteurs à Boorse sur l'objectivité du concept de santé mentale. Étant donné que cette caractérisation est partiellement erronée, je souhaite présenter les éléments qui permettront de montrer en quoi leurs positions peuvent être dites similaires et en quoi elles peuvent être dites différentes. Les positions de ces deux auteurs sont riches et complexes. Par souci de brièveté, je vais m'attarder principalement à leur position quant au caractère mythique du concept de maladie

mentale, puisque c'est en cela qu'il est intéressant de comparer leur position au naturalisme de Boorse¹.

En 1960, Szasz publie *The Myth of Mental Illness*, article célèbre dans lequel il défend l'idée que la maladie mentale n'existe pas en un sens scientifique, et qu'il s'agirait seulement d'un mythe. Dans cet article, Szasz s'oppose à l'idée selon laquelle les maladies mentales seraient tout autant réelles et objectives que les maladies somatiques. Selon Szasz, à la différence de la maladie somatique, qui réfère à une dysfonction physiologique, la notion de maladie mentale désignerait seulement les problèmes que rencontrent les individus dans leur vie, qu'ils concernent leurs besoins ou leurs valeurs personnels, leurs aspirations, leurs opinions, etc. :

I submit that the idea of mental illness is now being put to work to obscure certain difficulties which at present may be inherent – not that they need be unmodifiable – in the social intercourse of persons. If this is true, the concept functions as a disguise; for instead of calling attention to conflicting human needs, aspirations, and values, the notion of mental illness provides an amoral and impersonal "thing" (an "illness") as an explanation for problems in living (Szasz, 1959) [...] Mental illness exists or is "real" in exactly the same sense in which witches existed or were "real." (Szasz, 1960, p. 117)

Ainsi, selon Szasz, ce qui permet de déterminer ce qui constitue une maladie mentale est la déviation par rapport à une norme psychosociale, éthique ou légale (Szasz, 1960, p. 115). Le diagnostic de maladie mentale ne reposerait sur aucune base médicale, contrairement aux maladies physiques, qui elles reposeraient sur l'identification d'une

¹ La richesse de cette période historique majeure pour la psychiatrie ne peut être exposée en détail ici. Pour une analyse historique plus complète de la confrontation entre la psychiatrie et le mouvement antipsychiatrique, et de la diversité des positions dans ce mouvement critique, voir Sedgwick (1982), Bracken et Thomas (2010), Double (2006). Qui plus est, je ne peux pas rendre justice à la vision détaillée de Szasz et Sarbin dans ce chapitre. Pour une analyse plus détaillée de leurs théories respectives, voir Vatz et Weinberg (1994), Scheibe et Barrett (2016).

dysfonction physiologique. L'utilisation du concept de maladie mentale pour référer aux problèmes de vie, en plus d'être inutile, participerait à entretenir la confusion entre trouble physiologique et désir d'atteindre une vie idéale, ou comme le dit Szasz, entre une dysfonction et « a continuous struggle, not for biological survival, but for a « place in the sun, » « peace of mind », or some other human value. » (Szasz, 1960, p. 118). Szasz défend donc l'élimination des concepts de santé et de maladie mentale, parce que leur sens médico-scientifique serait illusoire.

Bien que la position de Sarbin se distingue de celle de Szasz à plusieurs égards², les deux auteurs font front commun quant au rejet de l'existence de la maladie mentale comme une catégorie scientifique valide. L'analyse de Sarbin est avant tout historique, en ce qu'il décrit la transformation du sens métaphorique initialement donné au concept de maladie mentale vers une signification mythique (Sarbin, 1967). Selon Sarbin, le concept de maladie mentale aurait d'abord été utilisé comme une métaphore. L'utilisation du concept de maladie pour référer à des désordres de conduite remonterait au 16^e siècle, où des comportements conçus comme « démoniaques » auraient peu à peu été conceptualisés comme des symptômes de causes naturelles, similaires aux maladies somatiques. Or, le sens métaphorique de la maladie mentale se serait peu à peu transformé en un sens littéral, où il serait assumé que des symptômes de maladie mentale seraient la manifestation d'une pathologie interne sous-jacente. Pour Sarbin, ce sens littéral donné au concept de maladie mentale serait illégitime étant donné que de telles pathologies, au sens de dysfonctions, ne sont que très rarement associées aux symptômes de maladie mentale. C'est pourquoi Sarbin parle du

² Une différence réside dans le fait que Sarbin s'oppose à l'utilisation du concept de maladie mentale, car elle sous-tendrait une conception dualiste forte de l'esprit et du corps, intenable à son avis. Dans l'article de 1976, Boorse critiquera aussi cet argument, mais je n'insisterai pas sur cet aspect ici, puisque je m'intéresse davantage à la thèse de l'absence d'un concept valide de trouble mental et à celle d'éliminativisme défendues par les auteurs.

« mythe » de la maladie mentale et défend l'élimination de ce terme (Sarbin, 1967, p. 451).

Il faut donc retenir qu'à la fois Szasz et Sarbin défendent 1) l'absence d'un concept valide de maladie mentale au sein de la psychiatrie et 2) un éliminativisme quant aux concepts de santé et de maladie mentales, et ce, en raison de l'argument 1. Pour les deux auteurs, le langage médical, avec les concepts de santé et de maladie qui lui sont propres, doit être réservé à la physiologie, la psychiatrie n'en faisant pas un usage légitime. C'est donc dans le cadre de ces critiques de la psychiatrie que Boorse rédige ses articles des années 1970 et défend la BST comme présentée dans la section 1.1. Dans *What a Theory of Mental Health Should Be?*, Boorse annonce trois objectifs qu'il entend poursuivre : 1) Défendre la possibilité d'un concept théorique de santé mentale, 2) Soutenir la thèse de l'absence d'un concept théorique de santé mentale au sein des théories acceptées de l'époque et 3) Suggérer que certaines approches cliniques de la psychiatrie, telles que la psychanalyse, pourraient contenir l'embryon d'une théorie de la santé mentale adéquate, car basées sur une théorie du fonctionnement mental normal. Alors que les points 1 et 3 sont propres à la position de Boorse, la thèse 2 est partagée par Boorse, Sarbin et Szasz. Boorse se distingue donc de Szasz et Sarbin en ce qu'il rejette l'approche éliminativiste face aux concepts de santé et de maladie mentales, mais tous les trois parviennent à la même conclusion suite à une analyse du concept de santé mentale employé par la psychiatrie de l'époque.

1.2.2 Que devrait être une théorie de la santé mentale?

Comme je l'ai déjà dit, Boorse est intéressé par ce que pourrait être une théorie de la santé mentale qui s'inscrirait légitimement dans le cadre du modèle médical. Pour Boorse, santé mentale et santé physique font partie d'un même concept général qui est celui de santé (Boorse, 1975, p. 50). Comme mis en lumière dans la section 1.1, plusieurs éléments clés composent la BST. Boorse fait émerger trois points qui doivent être retrouvés au sein d'une théorie de la santé, qu'elle concerne le physique ou le mental :

- 1) Le concept de santé doit être trouvé dans une discipline théorique (pour la santé somatique, il s'agit de la physiologie, voir la section 1.1.3).
- 2) Le concept de santé doit inclure la notion de fonctionnement normal (voir la section 1.1.1).
- 3) Ce qui constitue le fonctionnement normal doit être identifié sans jugement de valeur (je me concentre sur ce point au chapitre suivant).

Le point 2, le lien entre santé et fonctionnement normal typique, joue pour Boorse un rôle crucial dans une théorie médicale de la santé mentale. Selon Boorse :

Physical health is simply the special case obtained by focusing on the functions of physiological processes. Mental health, then, would be the special case obtained by focusing on the functions of mental processes; and so there is such a thing as mental health if there are mental functions (Boorse, 1976, p. 63)

Ainsi, la santé mentale théorique est, comme pour la physiologie, le fonctionnement normal par rapport au *design* de la classe de référence à laquelle appartient l'individu,

mais du point de vue psychologique. Le trouble mental est alors une déviation de la norme de fonctionnement psychologique de la classe de référence à laquelle l'individu appartient, conformément à ce que décrit la BST.

Étant donné que la santé est le fonctionnement normal, la possibilité de parler de santé mentale repose donc sur la possibilité d'identifier des fonctions mentales au même titre que nous identifions des fonctions physiologiques. Selon Boorse, deux conditions doivent être remplies pour que l'on puisse parler de fonctions mentales :

1. Des processus mentaux doivent jouer un rôle causal dans l'action.
2. Ces processus doivent se manifester selon une uniformité suffisante au sein de l'espèce pour pouvoir identifier des fonctions mentales typiques.

Boorse assume que ces deux conditions peuvent être remplies, et qu'en ce sens, il serait tout à fait légitime de parler de fonctions mentales. La première condition est peu discutée par Boorse. Boorse endosse rapidement les thèses défendues par Donald Davidson en rapport avec la causation mentale. Boorse pose sans grande discussion qu'il est d'accord avec l'idée de Davidson selon laquelle les arguments contre la causation mentale sont faibles. Boorse assume donc la thèse de la causation mentale comme probablement vraie. Sans m'engager dans une analyse approfondie de cette thèse mise de l'avant par Boorse, je tiens à souligner que cette prise de position permet de mettre l'accent sur le fait que Boorse ne défend pas un réductionnisme du mental au physique. En effet, Boorse soutient la spécificité de la psychiatrie, en ce qu'elle étudie les phénomènes mentaux plutôt que physiques. Selon Boorse, c'est une question empirique de savoir si les troubles mentaux peuvent être réductibles à des troubles neurologiques. Il serait tout à fait possible d'imaginer des troubles mentaux purement psychologiques. Je ne m'attarderai pas davantage aux enjeux entourant le dualisme mental/physique associé à la conception boorséenne de la santé mentale étant donné que je m'intéresse davantage à la seconde condition, celle de l'identification des

fonctions mentales normales³. Il convient simplement de retenir que Boorse n'a jamais défendu un réductionnisme du mental au physique et qu'il laisse plutôt cette question ouverte à des considérations empiriques. Naturalisme et réductionnisme ne vont donc pas forcément de pair pour Boorse (comme ils ne vont pas non plus de pair en philosophie de l'esprit; Fodor, 1974).

La seconde condition, l'idée selon laquelle les processus mentaux qui contribuent à l'action doivent le faire de manière assez uniforme au sein de l'espèce pour qu'il soit possible d'identifier un *design* fonctionnel mental, est également possiblement remplie. Boorse écrit : « If certain types of mental processes perform standard functions in human behaviour, it is hard to see any obstacle to calling unnatural obstructions of these functions mental diseases, exactly as in the physiological case » (Boorse, 1976, p. 64). Selon Boorse, de nombreux auteurs ont déjà étudié des caractéristiques typiques de l'esprit humain (p. ex. : Chomsky, Piaget, Freud, Darwin), ce qui suggère la possibilité de parvenir à une théorie du fonctionnement mental normal. Malgré l'apparente plasticité de l'esprit humain, Boorse énumère la perception, l'intelligence, la mémoire, la motivation, le langage, la souffrance et l'anxiété en guise d'exemples possibles de fonctions normales de l'espèce. Pour lui, ces fonctions semblent pouvoir représenter les différentes fonctions probables d'une théorie du fonctionnement mental typique de l'espèce humaine, bien que cela demande à être démontré de manière empirique (je reviens sur la manière selon laquelle Boorse envisage de démontrer cette typicité de manière empirique à la section 4). Selon Boorse, la plausibilité de ces deux éléments, la causation mentale et l'identification d'une théorie du fonctionnement mental normal, entraînent qu'il est tout à fait possible, en principe, d'imaginer un concept théorique de santé mentale comme on reconnaît l'existence d'un concept de santé physiologique.

³ Pour une excellente analyse de cet enjeu et du dialogue qu'il implique avec certains auteurs du mouvement antipsychiatrique, voir Demazeux (2015).

1.2.3 La thèse boorséenne de l'absence d'un concept théorique de santé mentale dans la psychiatrie des années 1970

Jusqu'ici, les interprétations de la théorie boorséenne de la santé mentale sont généralement conformes aux thèses réellement défendues par Boorse. Celui-ci reprend en effet les éléments constitutifs de la BST pour concevoir ce que serait un concept théorique de santé mentale. Cependant, il est important de noter que la thèse présentée dans la section précédente concerne uniquement la *possibilité* de parvenir à un concept théorique de santé mentale conforme aux exigences de la BST en physiologie. Elle ne concerne pas l'*existence* d'un concept théorique de santé mentale au sein de la psychiatrie des années 1970 et d'avant.

Pour bien comprendre cette nuance, il faut rappeler la manière dont Boorse parvient à sa définition du concept théorique de santé somatique et comment il procède pour analyser le concept de santé mentale. Une part de la confusion entourant les thèses réelles de Boorse peut probablement être expliquée par le fait que l'on a négligé cet aspect méthodologique. Comme expliqué à la section 1.1.3, Boorse s'inscrit dans l'approche du naturalisme méthodologique, qui requiert l'utilisation de la méthode de la reconstruction rationnelle pour parvenir à un concept théorique de santé mentale révélé par l'usage qu'en font les physiologistes. Or cette approche est avant tout descriptive. Par l'utilisation de la méthode de la reconstruction rationnelle, Boorse vise à reconstruire les critères d'utilisation des concepts théoriques de santé et de pathologie. Il ne les prescrit pas. Ainsi, tout comme Boorse analyse l'usage du terme de santé et de pathologie somatiques au sein de la discipline théorique qui lui est rattachée, soit la physiologie, il a aussi pour tâche de reconstruire l'usage des termes de santé et de trouble mental au sein des théories psychiatriques rattachées à la pratique psychiatrique. Or au terme de cette analyse du concept de santé mentale, les résultats

diffèrent de ceux de l'examen qu'il a mené pour la notion de santé somatique. Alors que pour la santé somatique, la reconstruction rationnelle réussit à fournir un concept théorique exempt de valeurs, pour la santé mentale, la reconstruction échoue :

It is quite likely that there is such a thing as mental health; yet the majority of "mental-health" theorists use methodologies that offer no assurance of finding out the first thing about it. They do not set out to investigate the normal functional organization of the human mind, as they ought to do if their subject is mental health (Boorse, 1976, p. 68).

En effet, selon Boorse, conformément à la BST, bien qu'il soit fort probable qu'il existe une telle chose que la santé mentale, la psychiatrie ne posséderait pas encore un concept théorique pour rendre compte de ce phénomène. Pour parvenir à cette conclusion, Boorse analyse la signification des concepts de santé et de trouble mentaux présents dans les théories psychiatriques de l'époque. Il dégage trois approches qui permettraient pour le moment aux théoriciens de la psychiatrie de parvenir à une conception de la santé mentale ainsi que les conceptions principales de la santé mentale résultant de ces méthodologies. Ces trois approches représenteraient selon Boorse les trois grandes conceptions de la santé mentale présentes dans les théories psychiatriques de l'époque. Ces trois conceptions seront ensuite confrontées aux exigences du modèle médical, afin de voir si elles font un usage légitime du concept de santé mentale.

Selon Boorse, au sein des théories psychiatriques principales de l'époque, la santé mentale serait comprise comme suit :

- 1) Une représentation d'un ensemble de traits de personnalité valorisés par la société;
- 2) Une représentation négative, comme ce qui ne serait pas une condition pathologique reconnue;
- 3) Une représentation de ce qui serait socialement acceptable.

Aucune de ces trois conceptions ne se présenterait comme un concept théorique de santé mentale valable en fonction du modèle médical. En ce qui a trait à la première conception, Boorse pose que cette conception serait obtenue par « affirmation de valeurs ». Les théoriciens choisiraient un ensemble de traits de caractère qui représenterait la personnalité idéale d'un être humain admirable. Il s'agirait alors de choisir les traits psychologiques qui sont socialement valorisés par la société, la culture ou le théoricien lui-même. Il apparaît évident que ce qui permet de parvenir à cette conception de la santé mentale est des décisions motivées par des valeurs, qu'elles soient sociales, culturelles ou personnelles, ce à quoi Boorse répond :

We are no more free to define mental health as the constellation of qualities we most admire than we would be for physical health. Mental health must be a constellation of qualities displayed in the standard functional organization of members of our species. Only empirical inquiry can show whether normal human beings have an even temper, engage in socially considerate behaviour, and advance the species-or make love with 'dignity and decency'. [...] The point is that a theory of health should be a description of how we are constituted and not how we would like to be. (Boorse, 1976, p. 70)

Ainsi, cette première voie, et la conception de la santé mentale à laquelle elle aboutit, ne fournit pas de concept théorique de santé mentale conforme aux exigences du modèle médical parce que la santé est définie en fonction de traits valorisés par la société ou le théoricien.

En ce qui concerne la conception 2, les théoriciens procéderaient par abstraction à partir de classes diagnostiques établies pour aboutir à une représentation négative de la santé mentale, comme ce qui ne serait pas une condition pathologique reconnue. L'exemple que donne Boorse est celui du théoricien psychanalyste L. Kubie qui, à partir de la catégorie diagnostique de névroses, propose une conception de la santé mentale comme ce qui ne serait pas des cas de névroses. La santé mentale désignerait alors des

conditions pour lesquelles il n'y aurait pas de conflit psychique. Selon Boorse, la difficulté principale que rencontre cette approche est le fait qu'elle présuppose que les conditions psychiatriques desquelles elle abstrait une conception négative de la santé mentale sont bel et bien des troubles mentaux. Le fait que certaines conditions soient associées à de la souffrance ou à un état d'incapacité, voire à une approche thérapeutique, ne permet pas, selon Boorse, de déterminer que cet état est pathologique :

However we may disvalue neurosis and seek to eradicate it, we cannot call it unhealthy until we know that the mind is not supposed to work that way. It is in no way obvious, and requires empirical support, that what clinicians see in their offices are usually cases of biological dysfunction. (Boorse, 1976, p. 71)

Cette route vers une conception de la santé mentale est donc également problématique en ce qu'elle présuppose que les catégories diagnostiques existantes sont bel et bien des troubles mentaux, sans connaissance empirique de ce que serait le fonctionnement mental normal. Cette conception ne renferme donc pas de concept théorique satisfaisant de santé mentale.

La dernière conception (3) aurait comme méthodologie les jugements sociaux du comportement. Portée par l'approche behavioriste, cette représentation conçoit le trouble mental comme un trouble du comportement. Parce que les behavioristes tentent d'éviter le recours aux processus mentaux dans leurs explications, les troubles du comportement sont dits pathologiques en ce qu'ils seraient des déviations sociales qui ne pourraient être expliquées au niveau physiologique : « Instead [the behaviourists] almost universally make a fatal error : they desert biological dysfunction in favour of social deviance. » (Boorse, 1976, p. 72-73). À cet argument, Boorse répond : « When, therefore, psychiatric theorists discard this biological constitution altogether in favour of the norms of a 'social and cultural system', it seems fair to say that they have stopped

talking about health. » (Boorse, 1976, p. 73). La conception selon laquelle la santé mentale correspond aux comportements socialement acceptables ne présente donc pas de concept théorique de la santé mentale, puisque celle-ci définit explicitement le trouble mental en fonction des valeurs de la société.

Ainsi, il apparaît clair que selon Boorse, les propositions 1, 2 et 3 ne sont pas des concepts théoriques de santé mentale qui satisfont les exigences de la BST. Bien que ces trois approches de la santé mentale présentent leurs défauts propres, le problème qu'elles rencontrent toutes est qu'elles ne reposent pas sur une théorie du fonctionnement mental normal, élément central de la BST. À ce sujet, Boorse écrit : « Apart from a theory of the structure and functions of the human mind, virtually all assertions about mental health are either misuses of language or flatly conjectural. » (Boorse, 1976, p. 81). L'argument de Boorse, qui s'applique pour les trois grandes conceptions de la santé mentale énumérées, peut être reconstruit de la manière suivante :

- A. D'un point de vue théorique, la santé est le fonctionnement d'un individu conformément au *design* fonctionnel de sa classe de référence.
- B. Or, on ne trouve pas dans les théories psychiatriques actuelles de théorie du fonctionnement mental humain.
- C. Donc, la psychiatrie actuelle ne possède pas de concept théorique de santé mentale.

En effet, les conceptions 1, 2 et 3 font reposer l'identification d'un état pathologique sur une norme de fonctionnement sociale, culturelle ou subjective, et non pas sur une norme de fonctionnement physiologique. Or l'analyse que Boorse a menée du concept de santé somatique, présentée à la section 1.1, a révélé que la connaissance du *design* fonctionnel est ce qui permet de distinguer un état sain d'un état pathologique. Les théories psychiatriques ne faisant pas référence à ce *design* fonctionnel mental, elles ne

possèdent pas de concept théorique de santé mentale. Ainsi, d'un point de vue descriptif, Boorse nie l'existence d'un concept théorique de santé mentale. Ainsi, bien que Boorse défende la *possibilité* de parvenir à un concept théorique de santé mentale, il ne défend pas l'*existence* de ce concept lorsqu'il utilise le naturalisme méthodologique, c'est-à-dire lorsqu'il reconstruit la signification de la santé mentale au sein des théories psychiatriques.

Par ailleurs, il convient d'attirer l'attention sur la proximité de la position de Boorse avec celles de Szasz et Sarbin. Bien que Boorse ait comme cible ces auteurs du mouvement antipsychiatrique, il faut bien voir en quoi les thèses de Boorse sont similaires à celles de Sarbin et Szasz quant à l'état actuel de la discipline psychiatrique. Boorse, Szasz et Sarbin nient tous les trois l'existence d'un concept théorique de santé mentale. Boorse reconnaît en effet que le critère pour déterminer si un état est pathologique relève actuellement du social ou du moral. Contrastant les résultats de l'analyse des concepts de santé somatique et mentale, Boorse écrit :

The doctrines of physiology about this functional design rest on empirical evidence and are either true or false. Where this factual component is wholly lacking, then, as in much psychological and psychiatric discussion, we do not have a health concept at all but one of moral acceptability or the good life for man (Boorse, 1976, p. 69).

En ce sens, le concept de santé mentale ne possède pas la validité nécessaire pour faire partie du vocabulaire médical. C'est exactement ce qui est défendu par Szasz et Sarbin, comme présenté dans la section 2.2. Ainsi, s'il faut catégoriser la position de Boorse en ce qui a trait à l'analyse du concept de santé mentale existant, Boorse peut être rangé du côté des normativistes, au même titre que Szasz et Sarbin. Or, Boorse s'écarte de ces deux auteurs quant aux implications de l'absence d'un concept théorique de santé mentale au sein de la psychiatrie, ce que je vais montrer dans la prochaine section.

1.2.4 En dehors du naturalisme méthodologique : l'attitude prescriptive de Boorse envers la psychiatrie

La section précédente a clarifié la thèse boorséenne selon laquelle la psychiatrie des années 1970 ne possède pas de concept théorique de santé mentale qui se conformerait à la BST. Or la position de Boorse ne s'arrête pas à ce volet négatif. Face à la conclusion selon laquelle la psychiatrie ne posséderait pas de théorie du fonctionnement mental normal, Boorse soumet la psychiatrie à un choix : « In the interests of conceptual clarity, psychiatry must either conform its usage of "health" to the functional paradigm or else devise new ways of speaking of the conditions with which it deals » (Boorse, 1976, p. 76). Ainsi Boorse offre deux possibilités à la psychiatrie : soit elle abandonne le vocabulaire médical, soit elle élabore une théorie du fonctionnement mental normal qui se conforme à la BST. Si la psychiatrie offre une théorie du fonctionnement mental normal, alors elle deviendra une discipline légitime du modèle médical et pourra alors maintenir son usage du vocabulaire médical. Si elle ne s'engage pas dans cette voie, alors elle devra renoncer à sa légitimité scientifique et à son affiliation médicale. Pour bien comprendre la position de Boorse, il convient de mettre au jour un présupposé derrière cette prescription : si la psychiatrie fait partie de la médecine, alors elle devrait se conformer, dans l'usage de ses concepts, aux exigences de ce modèle. Rappelons que pour Boorse, les concepts spécifiques de santé physiologique et de santé mentale appartiennent à un seul concept général de « santé ». Le concept de santé au sein du paradigme médical requiert la notion de fonctionnement normal, ce qu'a permis de mettre au jour la BST. La psychiatrie devrait ainsi tenter de faire correspondre ses concepts de santé et de trouble mentaux à ce critère définitionnel du concept général de santé théorique si elle veut faire partie de la médecine.

Au-delà de ces préoccupations théoriques et conceptuelles, l'impératif auquel Boorse soumet la psychiatrie s'explique également par des motivations sociales. Boorse soutient que l'utilisation des termes de santé et de pathologie sans base scientifique valide entretient la confusion et l'obscurité entourant le sens et la portée de ces concepts. Cette confusion pourrait avoir des implications sur le plan social. Discutant la troisième conception de la santé mentale présentée dans la section précédente, celle promue par les behavioristes, Boorse affirme :

To reject the straightforward extension of the health vocabulary to the psychological domain, while giving it instead a new meaning grossly disanalogous to its established use, is hardly a responsible terminological policy (Boorse, 1976, p. 73).

It is both dishonest and dangerous to throw the health concept out at the front door and then let it in at the back. If one wishes to avoid the theoretical inconveniences of the medical model, one must consistently relinquish its practical implications as well. There is no objection to advocating the use of techniques like behaviour modification on clinical populations. But if the aim of such techniques is to be 'adaptive behavior' rather than mental health, one should candidly admit that many of our assumptions about health – in particular, its desirability – may not carry over to adaptive behaviour at all (Boorse, 1976, p. 75).

Selon Boorse, le concept de santé est généralement associé, dans l'usage courant, à la désirabilité (Boorse, 1976, p. 61). Or, confondre la santé et la conformité sociale aurait pour conséquence de transférer le caractère désirable de la santé à la conformité, ce qui, pour Boorse, ne va pas du tout de soi. C'est pourquoi, à l'image de Szasz et de Sarbin, Boorse prescrit une attitude prudente et responsable quant à l'utilisation du vocabulaire médical dans le but d'éviter le contrôle et l'exclusion sociale.

Pour éviter cette confusion et ses implications sur le plan pratique, Boorse argue donc en faveur de l'élaboration d'une théorie du fonctionnement mental normal qui serait

conforme aux exigences de la BST. C'est ici que se trouve la différence majeure qui oppose Boorse à Szasz et Sarbin. Selon Boorse, la portée de la critique des antipsychiatres doit être limitée. L'absence d'une théorie du fonctionnement mental normal de l'esprit humain ne devrait pas entraîner l'abandon du concept de santé mentale, pour autant que la psychiatrie puisse fournir une telle théorie, ce que Boorse croit possible. La distinction entre les antipsychiatres et Boorse réside donc en ceci : les antipsychiatres ne croient pas possible de parvenir à un concept théorique de santé mentale. Szasz écrit :

My aim is more modest and yet also more ambitious. It is to suggest that the phenomena now called mental illnesses be looked at afresh and more simply, that they be removed from the category of illnesses, and that they be regarded as the expressions of man's struggle with the problem of how he should live. (Szasz 1960, p. 117)

Selon Szasz, parce qu'il n'existe pas de concept théorique de santé mentale en psychiatrie, la psychiatrie devrait renoncer à l'usage du vocabulaire médical. Or, Boorse croit que ce n'est pas parce que la reconstruction rationnelle du concept théorique de santé mentale échoue actuellement qu'il est impossible pour la psychiatrie de parvenir un jour à un tel concept :

And my final suggestion is that in spite of the methodological defects of current mental-health theory, it would be a mistake for the clinical disciplines to follow Szasz et al. in discarding mental health. Regardless of the origin of existing clinical criteria, there is reason to take many of them seriously as hypotheses about the normal human design. (Boorse, 1976, p. 76)

Autrement dit, alors que les antipsychiatres Szasz et Sarbin demandent à la psychiatrie de renoncer à toute prétention médico-scientifique en raison de l'absence actuelle de validité de son concept de santé mentale, Boorse croit que la psychiatrie devrait tendre vers le modèle médical grâce à l'élaboration d'une théorie du fonctionnement mental

normal. L'originalité de Boorse vis-à-vis Szasz et Sarbin est donc d'offrir à la psychiatrie un espoir de parvenir à intégrer de manière légitime le modèle médical et à user correctement de ses concepts, ce que Szasz et Sarbin lui refusent.

Le défi de Boorse est alors de justifier cet enthousiasme par rapport à la possibilité de parvenir à un modèle fonctionnel typique de l'esprit humain. Cet enthousiasme trouve sa source dans certains éléments théoriques existant déjà au sein de la psychanalyse freudienne.

1.2.5 La psychanalyse comme meilleure option disponible pour élaborer une théorie du fonctionnement mental normal

Selon Boorse, la théorie psychanalytique freudienne se présenterait comme la meilleure voie d'accès à un concept théorique de santé mentale conforme à la BST :

A well-developed example of the right kind of theory is already in the field. Formally speaking, psychoanalytic theory is the best account of mental health we have. It closely follows the physiological model by positing three mental substructures, the id, ego, and superego, and assigning fixed functions to each (1976, p. 78).

I am recommending psychoanalytic theory as our best model, apart from physiology, of what a theory of mental health should be. (1976, p. 78)

L'espoir que Boorse place dans la psychanalyse peut sembler surprenant à un lecteur contemporain étant donné sa baisse de popularité au cours des dernières décennies.

Plus encore, cet espoir est inattendu, car l'antipsychiatrie, les patients, le public et les psychiatres à orientation biologique critiquaient déjà la légitimité de la psychanalyse en tant que partie intégrante de la science et de la médecine au moment où Boorse écrivait ces lignes. En effet, au milieu des années 1970, la psychanalyse était largement marginalisée (Tsou, 2011). Par cet enthousiasme envers la psychanalyse, Boorse allait donc à l'encontre de l'évolution progressive de la psychiatrie de l'époque. Néanmoins, la foi de Boorse pourrait être démystifiée en se concentrant sur la façon dont la psychanalyse psychiatrique conçoit les troubles mentaux comme des dysfonctions. Parce que la psychanalyse comprend l'esprit en termes de sous-structures mentales (telles que le ça, l'égo et le surmoi dans la théorie de Freud), auxquelles sont assignées des fonctions mentales typiques, elle pourrait répondre aux exigences de la BST. Boorse encourage la psychanalyse parce qu'elle posséderait une théorie du fonctionnement mental normal⁴.

Or bien que la théorie psychanalytique soit prometteuse, elle présente deux importantes limites. La première est qu'elle confond parfois la santé mentale avec la personnalité idéale (Boorse, 1976, p. 79). Par exemple, certains théoriciens qualifient de normales des pathologies comme les névroses, ce qui entrerait en contradiction avec la BST. En effet, rappelons que selon la BST, la santé n'équivaut pas à un idéal de fonctionnement, mais seulement au fonctionnement typique de la classe de référence, selon un niveau d'efficacité statistiquement typique. La seule exception est, selon Boorse, des pathologies universelles induites par un environnement préjudiciable (p. ex. : les caries dans un environnement qui offre des options alimentaires très sucrées ou les irritations

⁴ Dans ce chapitre, je n'ai pas pour objectif de proposer une analyse exhaustive de la théorie psychanalytique. Mon but est plus modeste : je veux mettre en lumière les raisons pour lesquelles Boorse avait de l'espoir dans la psychanalyse dans les années 1970. Pour une analyse plus approfondie de la théorie, de son essor et de son déclin, voir Grob (1991), Wilson (1993), Hale (1995), voir également Schramme (2010, p. 42-43) pour une discussion critique de la recommandation de Boorse de se tourner vers la théorie psychanalytique du fonctionnement normal.

des poumons dans des environnements pollués, voir Boorse, 1977, p. 566, 1997, p. 86 pour une discussion). Si les névroses ne sont pas des pathologies induites par l'environnement, mais constituent un phénomène normal du développement de l'esprit humain, alors elles ne pourraient pas être qualifiées de pathologiques.

La seconde limite que rencontre la théorie psychanalytique est que les fonctions mentales qu'elle identifie ne sont pas encore pleinement validées :

These arguments show that some elementary functional claims about the species, together with a small body of widely accepted descriptive information about the mental processes of psychotics and neurotics, may give good grounds for provisionally calling these conditions pathological. Any stronger vindication of current clinical categories would require a detailed and well-confirmed theory of the functions of a normal human mind. (Boorse, 1976, p. 77)

Il faut donc insister sur le fait que même si Boorse pose que la théorie psychanalytique est la plus à même de révéler un concept théorique de santé mentale, elle n'y parvient pas véritablement. La thèse de l'absence d'un concept théorique de santé mentale au sein de la psychiatrie de l'époque s'applique encore ici, malgré le fait que Boorse suggère de pousser plus avant les recherches en s'inscrivant dans la théorie psychanalytique freudienne. Ce que Boorse juge valable de la théorie psychanalytique est donc uniquement qu'elle se présente sous la forme d'une théorie fonctionnelle de l'esprit humain. En cela, elle est la théorie qui fournit le plus de promesses d'une théorie du fonctionnement mental normal, et donc d'un concept théorique de santé mentale. Or étant donné que le modèle de la psychanalyse n'est pas encore pleinement validé, elle ne fournit pas encore ce concept.

Pour résumer les sections 1.2.1 à 1.2.5, j'ai cherché à clarifier la position de Boorse quant au concept de santé mentale. J'ai souhaité mettre en évidence le fait que Boorse

défend trois thèses principales quant à la définition de la santé mentale : la thèse de l'absence d'un concept théorique de santé mentale dans les théories psychiatriques des années 1970, la thèse de la possibilité pour la psychiatrie de parvenir à un concept théorique de santé mentale et la thèse de la désirabilité de parvenir à un tel concept. Dans la section suivante, je vais explorer la possibilité que la psychiatrie contemporaine possède un concept théorique de santé mentale au sein de son outil de classification officielle, le DSM.

1.2.6 La BST appliquée au DSM

Dans les sections précédentes, j'ai montré comment la BST de Boorse, lorsqu'appliquée au concept de santé mentale dans les années 1970, concernait la possibilité de parvenir à un concept objectif de santé mentale, et non son existence effective en psychiatrie. Il convient toutefois de se demander si les développements qu'a connus la psychiatrie occidentale contemporaine à travers les différentes éditions du DSM affectent les thèses de Boorse quant à la psychiatrie. Le choix de m'intéresser au DSM dans cette section est motivé par la méthode employée par Boorse. Rappelons ce que j'ai présenté à la section 1.1.3 : conformément à son naturalisme méthodologique, Boorse utilise généralement des manuels de classification pour reconstruire rationnellement la définition du concept théorique de santé. Le DSM étant l'outil officiel de classification des troubles mentaux en Amérique du Nord, il me paraît justifié de m'intéresser à un tel manuel. Ma question de recherche dans cette section est donc la suivante : est-ce que Boorse trouverait satisfaisante la conception de la santé mentale dans le DSM, spécifiquement au regard des exigences de la BST? À première vue, on pourrait être tenté de répondre par l'affirmative étant donné qu'à partir de la

troisième édition du DSM (1980), une définition formelle incluant la notion de dysfonction est présente. Or je vais arguer dans cette section que la réponse, aujourd'hui encore, est négative, et ce, en raison du caractère a-théorique du DSM depuis sa troisième édition. Pour le dire rapidement, dans la lignée de l'approche a-théorique du DSM, la définition du trouble mental adoptée par le DSM est elle-même a-théorique, c'est-à-dire sans attache par rapport à une théorie particulière des troubles mentaux (p. ex. : neurologique, psychanalytique, psychologique, comportementale, etc.). Pour cette raison, elle ne répond donc pas aux exigences de la BST.

Alors que le cadre conceptuel des deux premières éditions du DSM (APA, 1952, 1968) était de nature psychanalytique, la publication du DSM-III (APA, 1980) est marquée par l'adoption d'une approche dite « a-théorique et descriptive ». Alors que la psychiatrie des années 1960 et 1970 prenait un tournant scientifique, marqué par l'avancement de recherches en neurologie, en génétique et en pharmacologie, le DSM-III apparaissait comme fondamentalement a-théorique (voir Demazeux et Singy, 2015, p. xiv-xv pour l'idée selon laquelle le caractère a-théorique est paradoxal dans le contexte historique et scientifique dans lequel il est développé). L'approche a-théorique du DSM est à comprendre au sens où aucune présupposition n'est faite par rapport à l'étiologie sous-jacente à la manifestation des troubles mentaux (p. ex. : Spitzer et Williams, 1982)⁵. La classification psychiatrique se présente alors sous une forme purement descriptive, où les troubles mentaux sont caractérisés pour un ensemble de critères diagnostiques observables en clinique, principalement de nature comportementale (p. ex. : Aragona, 2009; Tsou, 2015). Ce changement de cadre conceptuel au sein du DSM s'explique par la volonté des concepteurs du DSM-III de se distancer de

⁵ Certains auteurs ont cependant remis en question l'idée que le système de classification du DSM est véritablement « purement a-théorique ». À ce sujet, voir Poland et coll. 2014; Whooley and Horwitz 2013. Par exemple, Whooley et Horwitz (2013) défendent que l'une des présuppositions théoriques du DSM est le réductionnisme biomédical, où les troubles mentaux sont conçus comme des troubles neurophysiologiques internes à l'individu.

l'approche psychanalytique, considérée par plusieurs comme spéculative, mais également d'assurer l'utilisation du manuel pour les psychiatres d'allégeance psychanalytique. L'idée était alors d'assurer l'acceptation du manuel par l'ensemble de la communauté psychiatrique (p. ex. : Demazeux et Singy, p. xv).

Bien que l'approche a-théorique du DSM ait été déjà critiquée par plusieurs (p. ex. : Hjørland, 2016; Poland *et al.*, 2014; Tsou, 2015), je veux m'intéresser aux implications de cette approche sur le concept de santé mentale promu par le manuel. Comme expliqué dans l'introduction, le DSM présente une définition formelle du trouble mental à partir de sa troisième édition. Cette définition formelle émerge pour trois raisons principales (Telles-Correia *et al.*, 2019) : 1) pour fournir une base conceptuelle à la classification a-théorique et empiriquement fondée du DSM, 2) pour répondre aux critiques provenant du mouvement antipsychiatrique qui prétendait que l'idée de trouble mental était un mythe et 3) pour tenter d'apaiser la crise entourant le diagnostic de l'homosexualité. Rappelons ici la définition formelle du trouble mental dans le DSM-III :

In DSM-III each of the mental disorders is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is typically associated with either a painful symptom (distress) or impairment in one or more areas of functioning (disability). In addition there is an inference that there is a behavioral, psychological, or biological dysfunction, and that the disturbance is not only in the relationship between the individual and society. (When the disturbance is not only in the relationship between the individual and society, this may represent social deviance, which may or may not be commendable, but is not itself a mental disorder.) (APA, 1980, p. 6)

Cette définition demeure à peu près inchangée dans les éditions ultérieures du DSM, mis à part dans le DSM-5. Dans la 5^e édition du DSM, la notion de dysfonction devient prioritaire et la notion de préjudice, un effet corrélé (voir Cooper, 2015 et Telles-

Correia, 2018 pour une analyse des modifications par rapport au critère de préjudice dans les différentes éditions). Ces définitions ont reçu beaucoup d'attention et de critiques (p. ex. : Cooper, 2015; Garson, 2007; Kinghorn, 2012; Varga, 2015, p. 114; Wakefield, 1992; Wakefield et First, 2003). Au regard du propos de cette section, la critique de Wakefield (1992) de la définition du trouble mental présentée dans le DSM-III-R (APA, 1987) est la plus parlante puisqu'elle concerne le caractère a-théorique celle-ci (voir aussi Rounsaville *et al.*, 2002, p. 3 pour une critique similaire). Je m'y attarderai donc dans ce qui suit.

Selon Wakefield, l'introduction de la notion de dysfonction dans la définition formelle du DSM est sous-théorisée. Étant donné que la notion de dysfonction n'est pas ancrée dans une théorie particulière de la fonction (p. ex. : la théorie évolutionniste de la fonction soutenue par Wakefield, 1992, 1999), elle devient vide de signification. Les éléments centraux de la définition du DSM sont donc réduits à l'ordre des préjudices encourus par les patients ou leurs proches. Ces préjudices sont à comprendre en termes de conséquences négatives provoquées par une condition particulière, et se traduisent par de la détresse ou un handicap. En raison de cette absence de définition de la notion de dysfonction, la définition du DSM devrait donc être reformulée comme suit : « A mental disorder is a mental condition that (a) causes distress or disability and (b) is not merely an expectable response to a particular event » (Wakefield, 1992, p. 236), laissant ainsi de côté toute référence à une théorie du fonctionnement mental normal. Comme le suggère Aragona, (2009, p. 3), l'absence de spécification de la notion de dysfonction s'explique par la volonté des concepteurs du DSM de demeurer agnostiques par rapport aux théories étiologiques sous-jacentes aux troubles mentaux du DSM, et ce, en vue de maximiser la possible utilisation des cliniciens formés dans des cadres théoriques différents. Qui plus est, la priorisation des critères relatifs au préjudice par rapport à la notion de dysfonction s'explique par la volonté de ne pas considérer l'homosexualité comme une pathologie mentale. Étant donné que l'homosexualité ne cause pas en elle-même de préjudices, elle ne peut pas être

considérée comme un trouble mental dans le DSM en fonction du critère de préjudice (Spitzer, 1973, 1981; Spitzer et Endicott, 1978; Spitzer et Williams, 1982; voir Bayer, 1980; Cooper, 2015; Telles-Correia *et al.*, 2018 pour une discussion). Mettre l'accent sur le caractère préjudiciable du trouble mental permettait aussi de répondre en partie aux préoccupations du mouvement antipsychiatrique : en attirant l'attention sur la détresse ou le handicap, les praticiens seraient davantage intéressés par les besoins réels des patients que par l'imposition de catégories biomédicales (p. ex. : Telles Correia *et al.*, 2019). Le fait que la notion de dysfonction devienne prioritaire dans la définition du trouble mental du DSM-5 semble s'éloigner des motivations initiales de la définition. Cette nouvelle conceptualisation peut par ailleurs entraîner des effets sur la manière dont les individus seront diagnostiqués avec des troubles mentaux dans l'avenir, voire mener à un surdiagnostic (voir Cooper, 2015 pour un exposé sur l'évolution de l'inclusion du critère de préjudice dans la définition et ses implications). Or en ce qui concerne la position de Boorse, cette modification n'a que peu d'implications. Voyons pourquoi.

Nous savons déjà que, selon l'interprétation de Boorse, la psychiatrie des années 1970 ne possède pas de concept théorique de santé mentale. Il convient désormais de tirer les implications d'une telle critique sur la position de Boorse quant au concept de trouble mental dans les éditions ultérieures du DSM. Reprenons les trois critères qui déterminent si un concept de santé, mentale ou physique, est théorique (section 1.2.2) :

- 1) Le concept de santé doit être trouvé dans une discipline théorique. Boorse exclut ici les concepts pratiques ou cliniques de la santé pour se concentrer sur le fondement théorique de ceux-ci.
- 2) Le concept de santé doit inclure la notion de fonctionnement normal.
- 3) Ce qui constitue le fonctionnement normal ne doit pas faire intervenir de jugement de valeur.

Dans un premier temps, il faut reconnaître que le DSM présente dans sa définition du trouble mental la notion de dysfonction, répondant ainsi au critère 2. Or comme je l'ai exposé précédemment grâce à la critique de Wakefield, cette notion ne s'ancre dans aucune discipline théorique particulière, manque d'ancrage qui la rend vide de signification. En ce sens, la définition du trouble mental du DSM ne répond pas au critère 1. Le concept de santé mentale n'est dépendant d'aucune théorie particulière en raison de la volonté des concepteurs du DSM de demeurer a-théorique pour une meilleure communication entre professionnels de la santé mentale. Il n'y a donc pas de théorie du fonctionnement mental normal sous-jacente au DSM. Quant au critère 3, il n'est pas rempli non plus. D'abord, étant donné que la notion de dysfonction est vide, il devient difficile de prétendre identifier le fonctionnement mental normal avec les méthodes des sciences naturelles. Plus encore, comme le montre Wakefield, il ne reste que les composantes de détresse ou de handicap à la définition du trouble mental, qui sont des notions évaluatives. La signification du trouble mental est donc dépendante de nos jugements de valeur, se distançant ainsi plus encore de la troisième exigence. Ainsi, il est possible de conclure que la psychiatrie contemporaine, avec la définition de trouble mental qu'elle présente depuis la troisième édition du DSM, ne possède pas de concept théorique de trouble mental. Or pour Boorse, posséder un tel concept est nécessaire pour faire partie du modèle médical. La psychiatrie contemporaine ne pourrait donc pas prétendre participer au modèle médical. En somme, en considérant les thèses soutenues par Boorse, il s'avère qu'il serait donc plutôt un critique du concept de trouble mental tel qu'il est décrit dans le DSM, et non pas un défenseur du concept neutre qui y figure.

Conclusion

Dans ce chapitre, j'ai exposé les grandes lignes de la BST comme défendue par Boorse. Je me suis intéressée à trois éléments clés de sa théorie, l'idée de fonction normale, la normalité statistique et la classe de référence, ainsi qu'au naturalisme méthodologique adopté par Boorse. Ce portrait général m'a permis de mettre en lumière la position spécifique de Boorse par rapport au concept de santé mentale. Je me suis attardée à un article spécifique de Boorse moins étudié dans les écrits, *What a Theory of Mental Health Should Be?* (1976), ce qui m'a permis de soutenir que les thèses de Boorse différaient quant au concept de santé somatique et de santé mentale. Mon analyse a des implications quant à la classification de Boorse comme un naturaliste et à l'opposition que l'on présente généralement entre Boorse d'une part, et les antipsychiatres comme Szasz et Sarbin d'autre part. Cette classification n'est juste que si elle concerne les thèses de la possibilité et de la désirabilité d'un concept théorique de santé mentale. La thèse de l'absence d'un concept de santé mentale en psychiatrie est défendue par tous les auteurs. En effet, la reconstruction rationnelle menée par Boorse aboutit aux mêmes résultats que l'analyse de Szasz et de Sarbin. Boorse, Szasz et Sarbin défendent donc tous les trois la thèse de l'absence d'un concept théorique de santé mentale au sein de la psychiatrie. Or Szasz et Sarbin défendent un éliminativisme quant au concept de santé mentale, tandis que Boorse incite la psychiatrie à élaborer ce concept par l'identification des fonctions mentales normales. Des nuances sont donc de mise lorsque ces positions sont contrastées.

Par ailleurs, j'ai montré que la position de Boorse quant au concept de santé mentale dans une approche naturaliste faisait de lui un critique de la psychiatrie contemporaine. Parce que le concept de trouble mental du DSM ne répond pas aux exigences de la BST, il ne s'agit pas d'un concept théorique. Il faudrait voir si d'autres domaines

théoriques permettent à la psychiatrie de fonder sa notion de dysfonction sur une théorie solide, comme la BST le requiert (c'est d'ailleurs d'une certaine façon ce qu'ont plaidé certains auteurs en promouvant l'adoption d'un cadre théorique particulier pour le DSM, p. ex. : Hyman, 2007; Kupfer, First et Regier 2002; Kupfer et Regier, 2011; Regier, 2008; Regier *et al.*, 2009;). Par ailleurs, le projet « RDoC » peut sembler une voie d'avenir dans le cadre de la BST. Or l'objet de ma thèse est la définition du trouble mental dans le DSM. Je laisse donc de côté cet aspect pour me concentrer pleinement au cadre du DSM. Dans le chapitre suivant, je vais explorer plus avant l'idée que le concept de santé est indépendant des valeurs, pour ensuite exposer les limites de la position naturaliste de Boorse.

Références du chapitre 1

- Aragona, M. (2009). The concept of mental disorder and the DSM-V. *Dialogues in Philosophy, Mental and Neuro Sciences*, 2(1), 1-14.
- Bayer, R. (1981). *Homosexuality and American psychiatry : The politics of diagnosis*. New York : Basic Books.
- Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & Public Affairs*, 5(1), 49-68.
- Boorse, C. (1976a). What a theory of mental health should be. *Journal for the Theory of Social Behaviour*, 6(1), 61-84.
- Boorse, C. (1976b). Wright on functions. *The Philosophical Review*, 85(1), 70-86.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44(4), 542-573.
- Boorse, C. (1997). A rebuttal on health. Dans J. M. Humber et R. F. Almeder (dir.), *What is disease?* (p. 1-134). Totowa (New Jersey) : Humana Press.
- Boorse, C. (2002). A rebuttal on functions. Dans A. R. Ariew, R. Cummins et M. Perlman (dir.), *Functions : New essays in the philosophy of psychology and biology* (p. 63-113). Oxford : Oxford University Press.
- Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, 39(6), 683-724.
- Bracken, P. et Thomas, P. (2010). From Szasz to Foucault : On the role of critical psychiatry. *Philosophy, Psychiatry, & Psychology*, 17(3), 219-228.
- Cooper, R. (2015). Must disorders cause harm? The changing stance of the DSM. Dans S. Demazeux et P. Singy (dir.), *The DSM-5 in perspective* (p. 83-96). Dordrecht : Springer.

- De Vreese, L. (2017). How to proceed in the disease concept debate? A pragmatic approach. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 42(4), 424-446.
- Double, D. B. (2006). Historical perspective on anti-psychiatry. Dans D. B. Double (dir.), *Critical psychiatry : The limit of madness* (p. 19-39). Londres : Palgrave Macmillan.
- Grob, G. N. (1991). Origins of *DSM-I* : A Study in Appearance and Reality. *American Journal of Psychiatry*, 148(4), 421-431.
- Hale, N. G. (1995). *The rise and crisis of psychoanalysis in America : Freud and the Americans, 1917-1985*. New York : Oxford University Press.
- Hjørland, B. (2016). The paradox of atheoretical classification. *Knowledge Organization*, 43(5), 313-323.
- Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, 8(9), 725-732.
- Kingma, E. (2013). Naturalist accounts of mental disorder. Dans K. W. M. Fulford, M. Davies, G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini et T. Thornton (dir.), *The Oxford handbook of philosophy and psychiatry* (p. 363-384). Oxford : Oxford University Press.
- Lemoine, M. (2013). Defining disease beyond conceptual analysis : An analysis of conceptual analysis in philosophy of medicine ». *Theoretical Medicine and Bioethics*, 34(4), 309-325.
- Lemoine, M. et Giroux, É. (2016). Is Boorse's biostatistical theory of health naturalistic? Dans Élodie Giroux (dir.), *Naturalism in the philosophy of health* (p. 19-38). s.l. (Suisse) : Springer International Publishing.
- Lilienfeld, S. O. et Marino, L. (1995). Mental disorder as a Roschian concept : A critique of Wakefield's « harmful dysfunction » analysis. *Journal of Abnormal Psychology*, 104(3), 411-420.

- Murphy, D. et Woolfolk, R. L. (2000a). Conceptual analysis versus scientific understanding : An assessment of Wakefield's folk psychiatry. *Philosophy, Psychiatry, & Psychology*, 7(4), 271-293.
- Murphy, D. (2005). *Psychiatry in the scientific image*. Cambridge (Mass.) : MIT Press.
- Nagel, E. et Hawkins, D. (1961). The structure of science. *American Journal of Physics*, 29(10), 716-716.
- Nordby, Halvor (2006). The analytic-synthetic distinction and conceptual analyses of basic health concepts. *Medicine, Health Care and Philosophy*, 9(2), 169-180.
- Poland, J., Von Eckardt, B. et Spaulding, W. (1994). Problems with the DSM approach to classifying psychopathology. Dans G. Graham et L. G. Stephens (dir.), *Philosophical psychopathology* (p. 235-260). Cambridge (Mass.) : MIT Press.
- Sadegh-Zadeh, K. (2000). Fuzzy health, illness, and disease. *Journal of Medicine and Philosophy*, 25(5), 605-638.
- Sadegh-Zadeh, K. (2008). The prototype resemblance theory of disease. *Journal of Medicine and Philosophy*, 33(2), 106-139.
- Sarbin, T. R. (1967). On the futility of the proposition that some people be labeled "mentally ill". *Journal of Consulting Psychology*, 31(5), 447-453.
- Scheibe, K. E. et Barrett, F. J. (2016). Sarbin's way : Overcoming mentalism and mechanism in psychology. *Theory & Psychology*, 26(4), 516-539.
- Schramme, T. (2010). Can we define mental disorder by using the criterion of mental dysfunction? *Theoretical Medicine and Bioethics*, 31(1), 35-47.
- Schwartz, P. H. (2004). An alternative to conceptual analysis in the function debate : Introduction ». *The Monist*, 87(1), 136-153.

- Sedgwick, P. (1982). *Psycho politics : Laing, Foucault, Goffman, Szasz, and the future of mass psychiatry*. Cambridge : Harper & Row.
- Sommerhoff, G. (1950). *Analytical biology*. Oxford : Oxford University Press.
- Spitzer, R. L. (1981). The diagnostic status of homosexuality in DSM-III : A reformulation of the issues. *The American Journal of Psychiatry*, 138(2), 210-215.
- Szasz, T. S. (1960). The myth of mental illness. *American psychologist*, 15(2), 113-118.
- Telles-Correia, D. (2018). Mental disorder : Are we moving away from distress and disability? *Journal of Evaluation in Clinical Practice*, 24(5), 973-977.
- Telles-Correia, D., Saraiva, S. et Gonçalves, J. (2018). Mental disorder : The need for an accurate definition. *Frontiers in Psychiatry*, 9(64). Récupéré de <https://www.frontiersin.org/articles/10.3389/fpsyt.2018.00064/full>.
- Tsou, J. (2011). The importance of history for philosophy of psychiatry : The case of the DSM and psychiatric classification. *Journal of the Philosophy of History*, 5(3), 446-470.
- Tsou, J. (2015). DSM-5 and psychiatry's second revolution : Descriptive vs. theoretical approaches to psychiatric classification. Dans S. Demazeux et P. Singy (dir.), *The DSM-5 in perspective* (p. 43-62). Dordrecht : Springer.
- Wakefield, J. C. (1992a). Disorder as harmful dysfunction : A conceptual critique of DSM-III-R's definition of mental disorder. *Psychological Review*, 99(2), 232-247.
- Wakefield, J. C. (1999). Disorder as a black box essentialist concept. *Journal of Abnormal Psychology*, 108, 465-472.
- Whooley, O. et Horwitz, A. V. (2013). The paradox of professional success : Grand ambition, furious resistance, and the derailment of the DSM-5 revision

process. Dans J. Paris et J. Phillips (dir.), *Making the DSM-5* (p. 75-92). New York : Springer.

CHAPITRE 2

OBJECTIVITÉ ET INDÉPENDANCE DES VALEURS DANS LA BST : LES CRITIQUES

Introduction

Dans le chapitre précédent, j'ai présenté les grandes lignes de la BST, la théorie naturaliste de Boorse. Dans les écrits en philosophie de la médecine et de la psychiatrie, cette théorie a été abondamment discutée et critiquée (p. ex. : Agich, 1983; Cooper, 2002, 2005; DeVito, 2000; Ereshefsky, 2009; Fulford, 2001; Giroux, 2009; Kingma, 2007, 2010, 2014; Varga, 2011; voir Ananth, 2008 pour une excellente recension de certaines de ces critiques). Plusieurs points problématiques y ont été relevés : le choix de la théorie de la fonction comme orientée vers des buts (OB), la méthode de l'analyse conceptuelle utilisée, l'exactitude de la description de la signification des concepts de santé et de pathologie dans la théorie médicale, l'impossibilité de Boorse à fournir une définition indépendante des valeurs, etc. Dans ce chapitre, je vais m'intéresser plus particulièrement à la prétendue neutralité de la BST vis-à-vis les valeurs sociales et culturelles.

Un premier défi qui se pose lorsque l'on se penche sur les prétentions de Boorse à fournir un concept objectif et neutre de pathologie est le sens exact des termes

« objectif » et « neutre ». Je vais donc d'abord éclaircir cette idée que les concepts de santé et de pathologie seraient objectifs, pour ensuite critiquer les prétentions de Boorse à fournir des définitions objectives de santé et de pathologie. Pour bien comprendre la notion d'objectivité employée par celui-ci, je vais analyser les arguments qu'il mobilise pour soutenir la thèse de l'objectivité des concepts de santé et de pathologie (section 2.1). Je vais ensuite examiner certaines des affirmations de Boorse qui peuvent sembler contredire cette thèse (section 2.2). Dans la dernière section (2.3), je vais critiquer les prétentions de la BST à fournir des concepts de santé et de pathologie indépendants des valeurs. Je terminerai le chapitre en soutenant qu'il n'est pas raisonnable de défendre que la BST soit véritablement indépendante des valeurs sociales et culturelles.

2.1 Les notions d'objectivité et de neutralité dans la BST

Comme je l'ai suggéré dans l'introduction, une source de l'ambiguïté de la BST réside dans la signification que donne Boorse aux notions d'objectivité, de neutralité et d'indépendance des valeurs. En effet, celui-ci définit l'objectivité comme ce qui serait neutre (Boorse, 1997, p. 56). Or la manière dont Boorse fait usage de ces notions, qu'il considère comme synonymes, semble variable et polysémique. Par ailleurs, il n'offre nulle part, à ma connaissance, d'explication détaillée et explicite de ce qu'il entend par ces termes. Cette ambiguïté est problématique si Boorse entend répondre aux critiques qui ont remis en question l'objectivité et la neutralité de sa définition. Je vais donc tenter de clarifier sa position vis-à-vis cet aspect des concepts de santé et de maladie, avant d'en présenter les critiques.

Pour bien comprendre la thèse selon laquelle les concepts de santé et de pathologie sont neutres (ou objectifs), il faut la situer dans le contexte du débat ayant lieu dans les années 1970, lors de la première élaboration de la BST. La position présentée par Boorse a pour objectif d'attaquer les approches normativistes prédominantes à cette époque. Boorse écrit :

More precisely, the orthodox view is that all judgments of health include value judgments as part of their meaning. To call a condition unhealthy is at least in part to condemn it; hence it is impossible to define health in nonevaluative terms. I shall refer to this orthodox view as normativism. (Boorse, 1975, p. 50)

Selon Boorse, une position normativiste « radicale » ou « forte » soutient que le concept de santé n'a pour contenu de signification que des jugements de valeur, sans signification descriptive, tandis qu'une approche normativiste « mixte » ou « faible » inclue à la fois des composantes évaluatives et descriptives (Boorse, 1975, p. 51). Ce qui lie les deux approches est la thèse selon laquelle les concepts de santé et de maladie ne peuvent être définis sans le recours à des jugements de valeur. Boorse s'oppose à ces deux types de normativisme en ce qu'il considère qu'un concept théorique de santé *n'implique pas* de jugements évaluatifs (Boorse, 2014, p. 684). La thèse selon laquelle la santé est un concept neutre ou objectif équivaut à soutenir que la définition de ce concept est indépendante des valeurs sociales ou culturelles.

C'est donc par opposition aux thèses normativistes faible *et* forte que Boorse situe son naturalisme. Il défend que les concepts théoriques de santé et de pathologie peuvent être définis en termes purement descriptifs, sans faire appel à des jugements de valeur. Ici, une précision est de mise : Boorse ne nie pas que des jugements de valeur interviennent dans la signification d'un concept général de la santé et de la pathologie partagée par les membres d'une société particulière. Rappelons que Boorse distingue le concept *théorique* de pathologie du concept *pratique* (voir le chapitre 1). Alors que

le concept théorique est purement descriptif, le concept pratique est influencé par des jugements de valeur reflétant des normes sociales et culturelles. L'influence des valeurs sociales et culturelles est donc à trouver dans le sens pratique des concepts de santé et de pathologie, et non pas dans son sens théorique. Cette distinction posée, Boorse affirme que le concept théorique de santé est indépendant des valeurs. Cette thèse est, selon moi, soutenue par au moins deux postulats qu'il convient de mettre en lumière : 1) la nature empirique des concepts de santé et de maladie et 2) l'indépendance de la science vis-à-vis les valeurs sociales et culturelles. Dans ce qui suit, j'exposerai brièvement ces deux postulats et j'offrirai une reconstruction des arguments qui permettent de mieux comprendre en quel sens Boorse défend l'indépendance des valeurs des concepts de santé et de pathologie, reconstruction qui s'inscrit dans une visée critique.

2.1.1 Premier argument : la nature empirique des concepts de santé et de pathologie

Selon Boorse, le concept de santé est indépendant des valeurs parce qu'il dépend d'abord et avant tout d'une connaissance des fonctions normales d'un organisme, qui peuvent être découvertes uniquement à l'aide d'une méthodologie scientifique. Plusieurs affirmations de Boorse vont en ce sens, par exemple :

[...] since the functional organization typical of a species is a biological fact, the concept of disease is value-free. Whether or not an organism is diseased *can be settled in principle by the methods of natural science*. (Boorse, 1976a, p. 63, les italiques sont les miennes)

Our conception of disease required no value judgment about what forms of human life are admirable or desirable. Diseases were interferences with an *empirically discoverable species design*. (Boorse, 1977, p. 571, les italiques sont les miennes)

Theoretical health, I argued, is the absence of disease; disease is only statistically species-subnormal biological part-function; *therefore, the classification of human states as healthy or diseased is an objective matter, to be read off the biological facts of nature without need of value judgments*. (Boorse, 1997, p. 4, les italiques sont les miennes)

Boorse affirme donc que nous pouvons accéder au *design* fonctionnel de la classe de référence par une méthodologie scientifique qui permet la collection de faits empiriques. Selon Boorse, en physiologie, cette méthode est statistique. Comme présenté au chapitre 1, Boorse soutient que c'est grâce à une idéalisation statistique que les physiologistes acquièrent la connaissance du *design* fonctionnel d'une classe de référence :

But the field naturalist abstracts from individual differences and from disease by *averaging* over a sufficiently large sample of the population. The species design that emerges is an *empirical ideal* which, I suggest, serves as the basis for health judgments in any species where we make such judgments. (Boorse, 1977, p. 557, les italiques sont les miennes; voir aussi Boorse, 1975, p. 58 pour une affirmation similaire)

Selon Boorse, ce *design* fonctionnel idéalisé statistiquement représente un fait de la nature plutôt qu'un jugement évaluatif basé sur des normes sociales ou morales.

Il peut être utile de reconstruire de manière plus explicite la thèse de Boorse dans une visée de clarté :

- 1) Le contenu empirique du concept de santé, soit le *design* fonctionnel de la classe de référence, est identifié par une méthode scientifique (la méthode statistique).
- 2) Les faits empiriques qui sont découverts par une méthode scientifique n'impliquent pas le recours à des valeurs sociales ou culturelles.
- 3) Ainsi, la signification théorique des concepts de santé et de pathologie ne fait pas appel aux valeurs sociales et culturelles.

Pour Boorse, l'utilisation de la méthode statistique, qui permet de mettre au jour le *design* fonctionnel de la classe de référence par idéalisation, est ce qui évite le recours à des jugements de valeur.

2.1.2 Deuxième argument : la neutralité de la science

Un autre argument dont dépend la position de Boorse est celui de la neutralité des concepts scientifiques. Celui-ci soutient que le concept de santé est neutre puisqu'il s'agit d'un concept *scientifique*. En effet, le concept théorique de santé que Boorse analyse est celui de la physiologie, une branche de la science. Pour Boorse, étant donné que la science est indépendante des valeurs et que les concepts de santé et de pathologie seraient scientifiques, ils seraient donc neutres. La nature scientifique des concepts serait en ce sens un gage de leur neutralité. Ces différentes citations illustrent ce deuxième volet argumentatif :

But behind this conceptual framework of medical practice stands an autonomous framework of medical theory, a body of doctrine that describes the functioning of a healthy body, classifies various deviations from such functioning as diseases, predicts their behavior under various forms of

treatment, etc. This theoretical corpus looks in every way *continuous with theory in biology and the other natural sciences, and I believe it to be value-free* (Boorse, 1975, p. 55-56, les italiques sont les miennes).

At least on cybernetic, etiological, or S&R accounts of function, the species design is a biological fact, making theoretical health judgments *as value-free as biological science itself is*. Theoretical health, on this view, belongs to the life-death family of biological concepts, not to the welfare-harm family of evaluative ones. (Boorse, 2011, p. 28, les italiques sont les miennes.)

Bien qu'à nouveau, ce raisonnement ne se présente pas de manière schématique dans les textes de Boorse, il peut être reconstruit comme suit :

- 1) Les concepts théoriques de santé et de pathologie sont des concepts scientifiques.
- 2) Si la science est neutre sur le plan des valeurs, alors ses concepts le sont également.
- 3) La science est neutre sur le plan des valeurs.
- 4) Les concepts de santé et de pathologie sont donc neutres sur le plan des valeurs.

Dans ce cas-ci, l'argument de Boorse repose sur le postulat selon lequel la science serait neutre sur le plan des valeurs. Je remets en question ce postulat dans le chapitre 4. Pour le moment, retenons donc que pour Boorse, la thèse de l'objectivité des concepts de santé et de pathologie équivaut à la thèse de la neutralité de ces concepts sur le plan de leur signification. Un concept objectif équivaut à un concept dont la signification est indépendante des valeurs sociales et culturelles. Cette thèse est soutenue par deux arguments, celui de la méthode scientifique employée pour parvenir à une connaissance de ces concepts, soit la méthode statistique, et celui de la nature scientifique des concepts de santé et de trouble.

2.2 Boorse contre Boorse : les valeurs admises dans la BST

Dans cette deuxième section, je pousserai l'analyse critique de la BST débutée au chapitre précédent en montrant que Boorse reconnaît lui-même la présence de valeurs à au moins deux moments au sein de son exposition des principes de la BST. Ces deux moments sont l'identification des buts de l'organisme et la détermination du seuil pathologique de fonctionnement. Notons que ces deux aspects ont déjà retenu l'attention au sein des écrits en philosophie de la médecine chez les critiques de Boorse (sur l'influence des valeurs sociales et culturelles pour identifier le seuil pathologique, voir Schwartz, 2007; sur l'influence des valeurs sociales et culturelles dans le choix des buts de l'organisme, voir De Vito, 2000 et Melander, 1997). Dans cette section, j'adopterai une approche différente en insistant sur le fait que Boorse lui-même reconnaît la présence de valeurs sociales et culturelles au sein de la BST, allant par là à l'encontre des thèses de neutralité de la BST. Mon analyse montrera que la position de Boorse au regard de la neutralité des concepts de santé et de pathologie est incohérente, puisqu'il admet lui-même l'influence des valeurs sur ces deux éléments de la BST. Étant donné que ces deux éléments jouent un rôle important dans la définition de la santé et de la maladie présentée par Boorse, montrer qu'il reconnaît que des valeurs influencent ces deux éléments peut aider à soutenir le rejet de la thèse de la neutralité de la BST.

2.2.1 Indétermination des buts de l'organisme et pluralisme des attributions fonctionnelles

Pour éclaircir la thèse de la neutralité des concepts de santé et de pathologie, il convient de traiter la manière dont Boorse considère la possibilité d'identifier les buts de l'organisme dans sa description des fonctions normales. J'ai expliqué dans le chapitre 1 que la notion de fonction défendue par Boorse était téléologique selon la conception de la fonction comme « orientée vers un but » (*goal-directed function*, ou OB). Dans ce cadre, les fonctions normales de l'organisme sont comprises comme les contributions causales statistiquement typiques des parties de l'organisme vers les buts de celui-ci, soit la survie et la reproduction. Rappelons également que pour Boorse, ces buts sont étudiés par la physiologie. Or, Boorse affirme également que les buts vers lesquels les systèmes tels que les organismes sont orientés sont en soi indéterminés. Ce sont les intérêts de recherche des scientifiques qui viendraient déterminer ces buts. Discutant de la téléologie des systèmes vivants, Boorse écrit :

To some extent, however, *these highest-level goals of organisms are indeterminate and must be determined by a biologist's interests*. It is a feature of the Sommerhoff analysis that whenever goal G causes G' within the range of environmental changes for which an organism is directed to G, that organism is also directed to G'. This phenomenon occurs constantly in biology. Most behavior of organisms contributes simultaneously to individual survival, individual reproductive competence, survival of the species, survival of the genes, ecological equilibrium, and so forth. *As a result, it appears that different subfields of biology (e.g., genetics and ecology) may use different goals as the focus of their function statements*. But it is only the subfield of physiology whose functions seem relevant to health. On the basis of what appears in physiology texts, I suggest that these functions are, specifically, contributions to individual survival and reproduction. This assumption has definite consequences for our health concept and should therefore be kept in mind.

Whatever goals are chosen, function statements will be value-free, since what makes a causal contribution to a biological goal is certainly an empirical matter. (Boorse, 1977, p. 556, les italiques sont les miennes)

Ainsi, c'est seulement lorsque les buts de l'organisme sont déterminés selon les intérêts de recherche des scientifiques que le concept de fonction peut être dit « indépendant des valeurs ». En effet, Boorse affirme qu'il serait possible de poser d'autres buts propres à l'organisme si notre contexte d'investigation scientifique était différent de la physiologie, comme c'est le cas de la biologie évolutionniste. Or selon Boorse, à la différence de la biologie évolutionniste, la physiologie s'intéresse aux fonctions qui participent à maintenir l'individu en vie et à maintenir ses capacités reproductives. De plus, la physiologie s'intéresse à la survie et à la reproduction actuelle de l'organisme, par contraste avec la biologie évolutionniste, qui utilise un référent temporel beaucoup plus grand. Biologie évolutionniste et physiologie peuvent donc viser des buts différents propres aux systèmes étudiés, sans que cela pose problème. Boorse semble donc défendre une approche pluraliste en ce qui concerne sa théorie de la fonction. Cette citation l'illustre :

This means that there is no sense in asking *which goal G is the goal at which the bird's behavior aims, and in respect to which its parts have functions, except in so far as this goal is clear from the context of discourse.* And the individual bird is only one system S. At the level of a population or species, new goal-directed processes seem to appear – for example, the maintenance of protective coloration in a changing environment and one can ascend still higher by viewing a whole ecosystem as a teleological unit [...] *Given this latitude for choice, what has happened is that various subfields of biology have carved out various slices of the teleological pie and interpreted their function statements accordingly.* In physiology the goal-directed system S is the individual organism and the relevant goals its own survival and reproduction. (Boorse, 1976b, p. 83-84, les italiques sont les miennes)

Ainsi, Boorse admet une pluralité d'attributions fonctionnelles possibles qui dépendent des contextes théoriques et des intérêts de recherche des scientifiques. Ce qui importe est d'être explicite quant au contexte théorique et de spécifier le système à l'étude, les buts de ce système et l'intervalle de temps considéré. Pour Boorse, si l'on explicite ces paramètres à la physiologie, on découvre que ses objectifs sont d'étudier les mécanismes typiques qui permettent à un organisme individuel (système) de se maintenir en vie et de se reproduire (buts de l'organisme) au présent (intervalle de temps).

Boorse affirme donc que l'identification des buts des organismes ainsi que les attributions fonctionnelles sont conditionnées par un ou des choix initiaux, qui relèvent des intérêts de recherche des scientifiques. Or, il ne considère pas que le choix des buts de l'organisme soit une menace à la thèse de l'indépendance des valeurs qui caractérise la BST. En effet, selon Boorse, sa conception de la santé et de la pathologie demeurerait neutre, malgré cette dépendance des buts de l'organisme à une décision quant au contexte d'investigation scientifique (Boorse, 2014, p. 693). Cette position a de quoi surprendre et a généré bon nombre de critiques à son encontre. Les critiques ont cependant visé différents points problématiques liés au choix des buts de l'organisme. Bien que la réponse de Boorse à ces critiques ne se présente pas de manière parfaitement schématique, il est possible de les regrouper en deux voies argumentatives qui répondent à deux critiques particulières :

- 1) Face à la critique de la normativité de la BST : une défense de la BST comme une conception *descriptive* du concept théorique de pathologie émanant de la méthode de la reconstruction rationnelle (section 2.2.2);
- 2) Face à la normativité des théories médicales : une défense du fait que ces choix n'impliquent pas que les concepts désignant ces phénomènes soient eux-mêmes « influencés par les valeurs » (section 2.2.3).

Je vais présenter chacune de ces deux critiques et la voie argumentative empruntée par Boorse pour y répondre, pour ensuite montrer la limite de ses réponses.

2.2.2 Une défense de la BST comme une conception *descriptive* du concept théorique de pathologie émanant de la méthode de la reconstruction rationnelle

Le premier angle de critique auquel Boorse a dû faire face est le suivant : parce qu'un système peut être dit posséder plusieurs buts biologiques, le choix fait par la BST des buts que sont la survie et la reproduction individuelles, par rapport à, par exemple, la survie de l'espèce, est normatif et n'est donc pas indépendant des valeurs (Boorse, 1997, p. 25). Dans « A Rebuttal on Health » (1997), Boorse tente de répondre aux critiques d'Engelhart (1976), d'Agich (1983) et de Brown (1985), qui l'auraient attaqué sur ce point. La réponse de Boorse consiste à dire qu'il ne s'agirait pas, pour la BST, d'un choix, mais simplement d'une description de ce qu'est la pathologie, comme le concept est compris par les physiologistes. L'objectif de Boorse est de décrire la science physiologique. L'enjeu est ici de savoir si Boorse propose une définition du concept de pathologie fidèle à l'usage qu'en font les pathologistes ou s'il propose une définition normative, qui demanderait à cet usage d'être modifiée. La stratégie argumentative de Boorse est alors de rappeler la nature descriptive de sa méthode qu'est la reconstruction rationnelle : « My answer, again, is that there is no choice here – that is simply what disease is, as the concept is best reconstructed from medical classifications. » (Boorse, 1997, p. 25). Autrement dit, la BST n'est qu'une reconstruction de la signification que donnent les physiologistes au concept théorique de pathologie. En cela, elle n'est coupable d'aucun choix étant donné qu'elle rend explicite, de manière descriptive, ce que les physiologistes entendent par

« pathologie ». La même réponse est donnée à Ereshefsky et De Vito (Boorse, 2014, p. 693). La BST demeurerait donc indépendante des valeurs parce qu'elle n'est pas normative sur ce que devrait être la signification du concept de pathologie : elle ne fait que reconstruire de manière descriptive les critères qui regroupent les différents cas de pathologie, critères implicites aux outils de classification en physiologie.

Cette réponse peut sauver la BST de cette première vague d'accusations, mais elle ne peut pas la sauver d'une critique qui viserait davantage la présence de valeurs au sein des théories scientifiques médicales. En effet, certaines critiques ne visent pas la fidélité de la description qu'offre la BST du concept de pathologie dans les théories médicales, mais la manière dont la médecine théorique elle-même parvient à son concept de pathologie. Si ce que décrit Boorse avec la BST est influencé par les valeurs, alors les concepts de santé et de pathologie ne peuvent pas être dits « indépendants des valeurs ». Dans ce débat, il est possible qu'une confusion survienne entre la normativité de la BST en tant que reconstruction rationnelle de la signification des concepts de santé et de pathologie et la normativité de la théorie médicale elle-même. Boorse y est pour quelque chose étant donné qu'il n'est pas rigoureux quant à ce qui peut être dit « indépendant des valeurs ». Or ce qui est clair, ce que la réponse qu'il a donnée à Engelhardt, la fidélité de la BST face à l'usage des concepts par les physiologistes, ne le sauve pas de cette deuxième attaque concernant les choix faits par les physiologistes pour délimiter le normal du pathologique. En effet, même si la BST est une description fidèle du concept théorique de pathologie utilisé par les physiologistes¹, si les physiologistes choisissent les buts du système à l'étude, il serait possible de soutenir que leur concept est guidé par des valeurs parce que provenant d'un choix et non pas d'un découpage « réel » de la nature indépendant de l'esprit humain. Le fait que les

¹ Lemoine et Giroux (2016) remettent en question le fait que la BST est une bonne analyse conceptuelle du concept théorique de pathologie parce qu'elle ne se base que sur la physiologie, excluant d'emblée d'autres disciplines biomédicales pertinentes à l'analyse de la signification du concept.

physiologistes décident des buts de l'organisme pourrait rendre le concept de pathologie normatif, même si la BST est elle-même une description non normative de ce processus par lequel les physiologistes décident des buts de l'organisme. Ainsi, même si Boorse a défendu que la BST était descriptive du concept théorique de santé, l'autre angle critique auquel il doit répondre est le suivant : la médecine, elle, est normative dans son choix des buts qu'elle décide d'étudier. La BST ne peut donc pas défendre que son concept de santé est indépendant des valeurs.

2.2.3 Les choix des buts de l'organisme

Boorse propose une autre voie de réponse qui pourrait venir pallier ce problème et qui consiste en ceci : une fois les buts de l'organisme choisis, l'étude des états pathologiques et de santé est objective, donc pas influencée par les valeurs. Dans « A Rebuttal on Health » (1997), Boorse discute le fait que la médecine ait choisi de reconnaître la reproduction comme un but indépendant de la survie de l'individu, bien que les fonctions reproductives puissent avoir un impact négatif sur les fonctions qui permettent de maintenir l'individu en vie (p. ex. : complications de grossesse comme les nausées et vomissements incapacitants, qui empêchent les femmes de s'alimenter dans les cas extrêmes). Ce choix de privilégier l'étude de la survie et de la reproduction de manière indépendante a été fait, de l'aveu de Boorse, pour des raisons évaluatives (Boorse, 1997, p. 28). Or, celui-ci maintient que cela n'impliquerait pas que le concept de pathologie contienne une ou des composantes évaluatives. Selon Boorse, l'intérêt que portent les scientifiques à un phénomène téléologique plutôt qu'à un autre n'impliquerait pas que le phénomène en question est lui-même évaluatif :

In sum, the choice-of-goals point does not make the BST normative unless one presupposes a form of normativism: that the concept of disease is defined by medical practice or medical ethics. That is in no way a necessary premise. Granted, physicians have a special professional commitment to treating disease. Similarly, teachers have a special professional commitment to spreading truth rather than falsehood. But that has no tendency to show that truth is a normative concept defined by educational practice. To choose wood over concrete to build your house with is an evaluative choice, but that does not make the concepts of wood and concrete value-laden (Boorse, 1997, p. 27).

La même voie argumentative est empruntée par Boorse dans « A Second Rebuttal on Health » (Boorse, 2014, p. 693-694)². Boorse semble y faire la distinction entre les choix qui ont trait aux phénomènes à étudier, relevant davantage des intérêts de recherche des scientifiques, et l'intrusion de valeurs au sein du concept une fois ces décisions prises.

Il convient de se demander si la réponse offerte par Boorse permet de bien répondre à cette deuxième vague de critiques. Selon mon interprétation, ces critiques ont cherché à montrer que le concept théorique de pathologie n'est pas neutre, parce que déterminé par un choix préalable à l'investigation scientifique. Il semble qu'ainsi construite, la réponse de Boorse est un premier pas vers une réponse satisfaisante, quoiqu'encore incomplète. En effet, Boorse n'a pas montré en quel sens le choix des buts de l'organisme par les physiologistes n'affectait pas l'indépendance du concept de fonction normale une fois le choix fait, et ainsi la neutralité des concepts de santé et de pathologie. En effet, un élément sous-jacent à la critique ayant visé Boorse, mais qui

² Il semble ici que Boorse confond la méthode d'investigation scientifique des phénomènes utilisée par les physiologistes pour identifier les buts de l'organisme, qui elle implique des choix, et la méthode d'analyse conceptuelle ou de reconstruction rationnelle qu'il emploie. Or, le fait que Boorse réussisse dans son analyse de la signification du concept de pathologie ne veut pas dire que les pathologistes peuvent identifier les cas de pathologie sans faire appel à des choix évaluatifs. La méthode de Boorse est certes descriptive (ou elle se veut comme telle), mais cela n'implique pas que la méthode scientifique le soit.

n'est pas abordé par celui-ci (peut-être parce que non explicité par les critiques elles-mêmes, si on veut être charitable envers Boorse), est l'idée que si les physiologistes choisissent de manière arbitraire ou selon des critères purement normatifs (subjectifs, sociaux, culturels, politiques) les buts de l'organisme, alors le concept de pathologie ne peut pas être neutre. En effet, le concept de fonction normale ne reposerait pas sur des bases uniquement empiriques. Si la médecine « plaque sur la nature » des buts comme la survie et la reproduction, qui sont simplement le reflet de normes sociales ou culturelles manifestant le fait que nos sociétés valorisent ces deux phénomènes, alors le concept de fonction biologique n'est pas parfaitement neutre. Le concept de fonction normale serait construit en référence à ces deux idéaux sociaux, reflétant la préférence de la société pour la survie et la reproduction individuelles. Parce que le concept de fonction normale serait guidé par des valeurs, le concept de pathologie ne pourrait pas être complètement indépendant des valeurs. La tâche centrale qui incombe à Boorse est alors de montrer que la physiologie choisit, parmi un ensemble de buts *réels* ou *naturels* des systèmes téléologiques, ceux qui l'intéressent. Bien que déterminés sur une base décisionnelle, les buts des systèmes biologiques ne seraient pas arbitrairement choisis, ou choisis sur la seule base des normes sociales. Cet aspect de réponse est manquant des textes de 1997 et de 2014.

Il faut se tourner vers « A Rebuttal on functions » (Boorse, 2002) pour trouver une défense de l'indépendance du concept de fonction biologique vis-à-vis les valeurs qui semble davantage prometteuse. Boorse y répond à la critique de Melander (1997), pour qui les buts des organismes seraient empiriquement indéterminés, et à celle de Schaffner (1993), pour qui le langage téléologique impliquerait nécessairement des jugements de valeur. Pour ces deux auteurs, l'identification des buts de l'organisme serait donc purement arbitraire ou évaluative. La réponse de Boorse n'est pas très détaillée, mais je vais tenter de faire émerger sa position en référant à ce texte ainsi qu'à l'article « Wright on Functions » (1976b), où il est plus explicite. Ce dernier semble soutenir qu'il ne serait pas arbitraire de poser la vie et la reproduction comme

les buts hiérarchiquement supérieurs d'un système. Autrement dit, les organismes tendent *naturellement* à la survie et la reproduction, indépendamment du fait que nos sociétés valorisent de tels phénomènes. S'inspirant de Sommerhoff (1950), Boorse donne une définition précise des buts d'un système biologique : un système réellement orienté vers des buts se reconnaît par la présence de mécanismes de compensation présents dans le système, qui maintiennent les buts en place lorsque surviennent des changements dans l'environnement : « To say that an action or process A is directed to the goal G is to say not only that A is what is required for G, but also that within some range of environmental variation A *would have been modified* in whatever way was required for G. » (1976b, p. 78) Et plus loin :

When a process appropriate to several ends at once has a true goal, I suggest it is because the process is produced by an internal mechanism which standardly guides pursuit of that goal but not the others. [...] With Sommerhoff as amended, then, one can say that organisms are centers of activity which is *objectively directed* at various goals – for example, survival and reproduction. (Boorse, 1976b, p. 79, les italiques sont les miennes)

Ces affirmations de Boorse suggèrent qu'au-delà de la reconstruction rationnelle de la signification des concepts de santé et de maladie, la BST propose aussi une description de la manière dont la physiologie acquiert la connaissance des buts de l'organisme. Pour Boorse, les buts sont « empiriquement donnés » (Boorse, 1975, p. 58). Ils ne sont donc pas déterminés arbitrairement. Ils peuvent être identifiés grâce à la présence de mécanismes de rétroaction (*feedback mechanisms*) qui maintiennent la poursuite de certains buts malgré des changements dans l'environnement. L'exemple utilisé par Boorse est celui des fonctions homéostatiques telles que la régulation de la température en fonction de changements d'environnement que subirait l'organisme : si un changement de température survient, des mécanismes de compensation seront activés pour maintenir la température de l'organisme, et ainsi le maintenir en vie (Boorse, 1997, p. 21). Or, ces mécanismes de rétroaction peuvent être trouvés à plusieurs

niveaux et découpés de plusieurs façons. Il revient donc à chaque discipline d'établir lesquels des systèmes elle veut étudier et quels buts des systèmes l'intéressent. Les disciplines scientifiques choisissent alors, relativement à leurs intérêts, parmi les divers buts des systèmes grâce à l'identification de mécanismes de rétroaction.

Il serait cependant difficile pour Boorse de soutenir que ce choix de la survie et de la reproduction est guidé par des raisons purement théoriques :

The idea that biological concepts are inherently value-laden seems to have one of two sources. First, it may rest on confusing our motive for studying something with the thing itself. Though we mostly study physiology in order to discover how the body works so that we can stay alive and well, that is equally true of, say, geology or atmospheric science. We study almost everything for our own benefit. Thus, our major reason for caring what creates and maintains planetary atmospheres is that we will die with the wrong one. But that fact has no tendency to show that planetology or meteorology is a science of value-laden propositions. (Boorse, 1997, p. 57)

Ainsi, et Boorse le reconnaît, c'est bien parce que la survie et la reproduction sont valorisées sur le plan social ou moral que ces buts sont étudiés par la physiologie. Par contre, il faut noter que généralement, le choix des phénomènes étudiés ne pose pas de problème particulier pour quiconque soutiendrait une vision de la science comme neutre (je développe davantage cet aspect au chapitre 4). Boorse soutient qu'une fois ce choix fait d'étudier certains buts comme la survie et la reproduction (qui sont des buts bien réels, identifiés par la présence de mécanismes de rétroaction), et que bien que ce choix soit justifié par des considérations « influencées par les valeurs », le phénomène téléologique étudié est, lui, indépendant des valeurs. Autrement dit, une fois que la reproduction est identifiée comme le but à étudier, les fonctions qui maintiennent la reproduction peuvent être dites indépendantes des valeurs.

Pour Boorse, le fait que l'identification des buts de l'organisme dépende de valeurs n'implique pas que ces choix soient arbitraires, au sens où un scientifique ne pourrait pas poser n'importe quel but comme un but « réel » de l'organisme dépendamment des valeurs de la société. Une fois le contexte d'investigation et les intérêts de recherche explicites, il est tout à fait possible de décrire les fonctions naturelles d'un système, en fonction des buts à étudier, déterminés par les intérêts de recherche, mais réellement présents dans la nature. En somme, la meilleure option de défense qui s'offre à Boorse contre ses critiques est de rappeler que le choix des buts n'est pas un choix arbitraire, ni un choix guidé uniquement par des valeurs sociales, culturelles ou politiques, mais un choix déterminé par l'observation de réels phénomènes téléologiques.

Malgré tout, cette réponse demeure encore insatisfaisante au regard des critiques qui lui ont été faites. Les critiques remettaient en question les prétentions de Boorse à proposer une définition du concept de pathologie parfaitement indépendante des valeurs. Même s'il défend l'idée que les buts sont « réels », il devrait aussi montrer que le choix d'étudier certains buts par rapport à d'autres ne conduit pas à abandonner la thèse de l'indépendance des valeurs du concept de pathologie. Cette défense est absente des textes de Boorse et me semble beaucoup plus difficile à formuler. En effet, celui-ci devrait plutôt admettre l'influence de valeurs à ce moment précis qu'est le choix des buts de l'organisme à l'étude. Dans la section suivante, je vais me pencher sur un autre endroit dans les théories de Boorse où les valeurs semblent s'immiscer : le choix du seuil pathologique de fonctionnement.

2.2.4 Le seuil pathologique et le problème de la démarcation

La section précédente concernait l'identification des fonctions normales par rapport aux buts d'un système. Cette identification des buts de l'organisme permet de distinguer les fonctions normales des autres traits et comportements d'un organisme, et par là permet de déterminer le *design* fonctionnel de la classe de référence, un élément central de la BST. Autrement dit, l'identification des buts de l'organisme permet de savoir que la circulation du sang est une fonction normale que doit remplir l'organisme parce que cette activité contribue causalement et de manière statistiquement typique aux buts de l'organisme, soit la survie et la reproduction. Selon la BST, pour être sain, un premier critère est donc qu'un organisme doit réaliser ses fonctions normales, typiques de sa classe de référence. Or selon la BST, ce n'est pas tout de réaliser cette fonction typique. Un second critère est que les parties de l'organisme doivent également réaliser cette fonction selon *une efficacité statistiquement typique*. Autrement dit, au-delà de l'identification de la circulation du sang comme la fonction normale du cœur, un autre critère qu'il faut préciser est le niveau d'efficacité normal de fonctionnement. Dans le cas du battement du cœur, ce niveau est généralement obtenu via une mesure de performance (dans ce cas-ci, la fraction d'éjection (FE), c'est-à-dire le pourcentage d'éjection du sang lors d'un battement). De l'aveu même de Boorse, ici aussi, certaines conventions de nature non empirique peuvent venir influencer le choix des niveaux de performance sain et pathologique. Ainsi, l'influence des valeurs peut se produire lorsque les pathologistes identifient ce seuil de performance de la fonction normale, seuil qui permet là aussi de démarquer le normal du pathologique. Cet aveu ouvre à nouveau la voie à la reconnaissance de la présence de valeurs au sein des concepts de santé et de pathologie.

Comme le dit Boorse, un état est dysfonctionnel si la réalisation de la fonction normale s'effectue selon un niveau de performance « très en dessous » (*far below*, Boorse, 1977, p. 559; voir aussi Boorse, 1997, p. 7-8) de la moyenne statistique de la classe de référence auquel l'organisme appartient. Le problème de la démarcation surgit alors : à quelle distance de la moyenne un état devient-il pathologique? Et comment peut-on déterminer cette distance? Le problème est de déterminer à quelle distance par rapport à la moyenne un état devient pathologique, ce que Schwartz (2007) a nommé le « problème de la démarcation » (*line-drawing problem*). En réponse à cette question, Boorse soutient que cette distance est déterminée de manière conventionnelle :

Abnormal functioning occurs when some function's efficiency falls more than a certain distance below the population mean. [...] [T]his distance can only be conventionally chosen, as in any application of statistical normality to a continuous distribution. The precise line between health and disease is usually academic, since most diseases involve functional deficits that are unusual by any reasonable standards. (Boorse, 1977, p. 559)

Bien que conventionnelle, la ligne serait tracée selon des « standards raisonnables » (Boorse, 1987, p. 371). Ainsi, selon Boorse, bien que les fonctions normales puissent être identifiées de manière neutre, la démarcation entre le niveau normal d'efficacité d'une fonction et le niveau pathologique est bel et bien déterminée en faisant appel à des conventions. Dans cette section, je présente l'argument de Schwartz qui tire les implications de cet aspect conventionnel de la BST sur les prétentions de Boorse à fournir un concept de pathologie indépendant des valeurs.

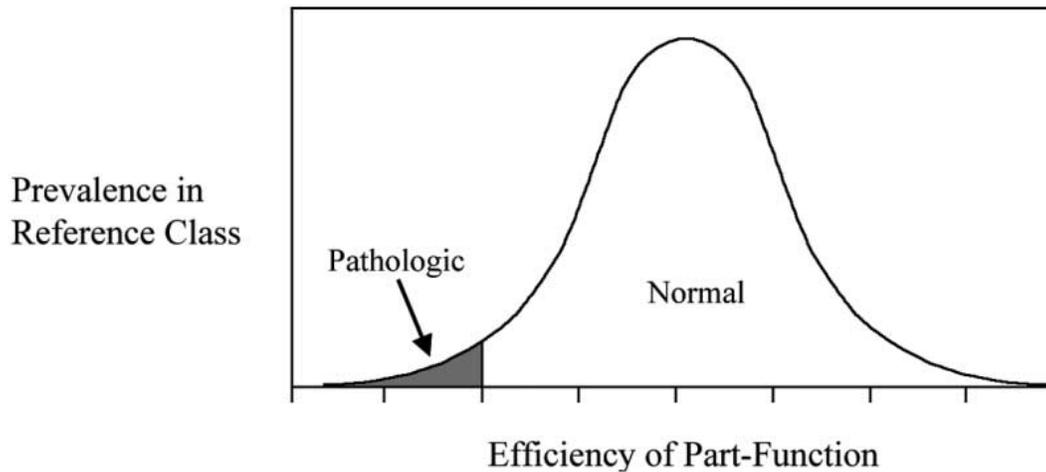


Figure 1. Courbe de la distribution de l'efficacité des fonctions réalisées

Source : Schwartz 2007, p. 373. Il s'agit d'un graphique repris d'articles de Boorse (1987, p. 370; 1997, p. 8), mais où les axes ont été modifiés conformément à l'intention initiale de Boorse.

Comme je l'ai présenté précédemment, selon Boorse, ce sont des mesures statistiques de la distribution du niveau d'efficacité de la réalisation d'une fonction normale qui détermine où il faut tracer la ligne entre le normal et le pathologique. L'endroit où tracer la ligne entre un niveau de performance peu élevé et un niveau de performance dysfonctionnel est établi en référence à des mesures statistiques qui représentent la distribution du niveau de performance de la fonction au sein d'une classe de référence. Plusieurs fonctions sont mesurées selon un continuum qui représente le niveau de performance de cette fonction, du niveau de fonctionnement normal vers le niveau de fonctionnement pathologique. Boorse propose que ce soit le 5 % inférieur d'une distribution statistique qui constitue la section pathologique de la courbe. Rappelons que pour Boorse, ce 5 % est conventionnel, voire « arbitraire » dû à la nature continue de la distribution (voir la figure 1). Or, cet aspect conventionnel ne veut pas dire que la ligne pourrait être tracée n'importe où sur le continuum. Par exemple, en médecine, le seuil pathologique est généralement tracé à partir de 2 standards de déviation sous la moyenne de la classe de référence. La ligne peut alors être tracée de manière arbitraire

tant et aussi longtemps qu'elle est tracée dans cette échelle de variation. Pour Boorse, l'endroit exact où il convient de tracer la ligne est « évident » la plupart des cas (Boorse, 1997, 2004).

Contre l'« évidence » plaidée par Boorse, Schwartz soutient qu'il n'est pas possible de déterminer l'endroit exact où tracer la ligne sur une distribution continue sans faire appel à un autre critère, celui des conséquences préjudiciables pour les individus généralement associées à ce pourcentage statistique. Pour faire émerger cette insuffisance de la BST, Schwartz présente deux cas qui posent problème à la théorie : les pathologies communes et les populations saines. Le problème des pathologies communes se présente dans des populations où 10, 20, voire 30 % et plus de la population présente une condition pathologique. Se référer au seuil de 5 % suggéré par Boorse, voire aux deux standards de déviation sous la moyenne, impliquerait de ne pas reconnaître des cas intuitifs de pathologie puisque le seuil pathologique serait beaucoup plus bas. Schwartz utilise l'exemple d'un homme de 70 ans dont le cœur aurait pour FE 20 %. Supposons qu'au sein de la classe de référence des hommes âgés de 65 à 80 ans, entre 10 et 20 % des hommes présentent une FE de 20 %. LA BST ne serait pas en mesure de qualifier ces états comme « pathologiques », étant donné que cette FE ne s'écarte « pas assez » de la performance normale (puisque pour Boorse, seulement les 5 % les plus bas peuvent être dits dans un état pathologique). Le problème des pathologies communes montre une limite de l'approche de Boorse : sans recours à d'autres critères que cette déviation statistique conventionnelle, il n'est pas possible de considérer une FE de 20 % comme pathologique si un trop grand nombre d'individus en est affecté dans la classe de référence. Bien que l'exemple vise une FE dysfonctionnelle, dans la classe de référence des personnes âgées, d'autres pathologies pourraient être « communes » et donc ne pas se voir attribuer le statut de dysfonction, ce qui apparaît problématique (p. ex. : les caries, l'arthrite, l'Alzheimer, etc.). Il s'agit donc d'une limite importante pour Boorse et sa BST.

Par contraste, le problème des populations saines se manifeste lorsqu'une population présente un taux de prévalence de pathologie très faible (2 % et moins) par rapport à une fonction donnée. Dans ce cas-ci, des personnes intuitivement en santé seraient considérées dans un état pathologique. Supposons une population d'hommes âgés de 20 à 30 ans. Si tous ces hommes ont une FE entre 50 et 70 %, ce qui est considéré comme sain, les hommes présentant la FE la plus basse, soit 50 %, seront considérés pathologiques. Or, leur cœur pompe le sang selon un niveau suffisant, et cette FE ne cause aucun symptôme incapacitant ni aucun risque particulier de développer des symptômes dans le futur. Il semble donc qu'avec le seul critère de la distance statistique, Boorse serait contraint de considérer des individus sains comme pathologiques. Il faut donc un autre critère pour expliquer pourquoi la ligne entre un niveau de performance bas et un niveau de performance dysfonctionnel est tracée, outre la distance statistique.

Ainsi, les problèmes des pathologies communes et des populations saines montrent que si l'on suivait la BST, la ligne entre un niveau de fonctionnement bas et pathologique serait tracée au mauvais endroit dans plusieurs cas. En effet, si aucune référence n'est faite aux conséquences négatives de ce seuil, évaluées en termes de symptômes préjudiciables pour les individus (p. ex. : développement de symptômes incapacitants, risque plus élevé de mortalité, etc.), on assiste à des résultats aberrants, où des pathologies évidentes ne pourraient pas être identifiées (problèmes cardiaques, Alzheimer), et des conditions saines seraient jugées pathologiques (FE de 50 % dans la classe de référence des hommes de 20 à 30 ans). Étant donné que plusieurs conditions correspondent aux problèmes des pathologies communes ou des populations saines, l'argument de Schwartz est d'autant plus convaincant³. D'autres auteurs ont d'ailleurs emboîté le pas à cette critique en montrant que le même problème surgissait pour

³ Pour une réponse à Schwartz et une proposition alternative, voir Hausman (2012, 2014).

d'autres conditions telles que le cancer ou la tuberculose (p. ex. : Rogers et Walker, 2017). Ces discussions montrent donc qu'au-delà du caractère conventionnel de la démarcation entre un niveau de performance bas et dysfonctionnel, ce sont des valeurs en lien avec les conséquences négatives possibles pour l'organisme qui viennent déterminer le seuil pathologique de fonctionnement. L'implication du problème de la démarcation est donc la présence de valeurs au sein de la BST, ici dans le cas de la démarcation.

2.3 Les classes de référence et l'indépendance des valeurs de la BST

L'ambition de Boorse de fournir un concept de pathologie qui ne fait pas intervenir les valeurs ne fait pas consensus et a été largement critiquée depuis la parution de la première élaboration de la BST dans les années 1970. Dans la section précédente, j'ai montré que Boorse reconnaissait la présence de valeur à au moins deux moments au sein de la BST : au moment du choix des buts de l'organisme à étudier et à celui de l'identification du seuil pathologique. Je vais désormais me concentrer sur les objections qui ont visé la neutralité de la BST dans son rapport avec la théorie médicale. Autrement dit, je vais analyser les arguments qui ont visé la manière dont les médecins parviennent à la connaissance de ce qui constitue un état normal et le rôle des valeurs dans ces processus. Comme présenté dans le chapitre précédent, chez Boorse, la normalité implique deux éléments : 1) l'identification des fonctions normales et 2) la détermination du niveau d'efficacité statistiquement normal de fonctionnement. Dans ce qui suit, je vais présenter trois arguments qui ont visé respectivement l'idée de classe de référence et son lien avec les éléments 1 ou 2. Ces trois arguments visent à démontrer l'influence des valeurs sur la manière dont les classes de référence sont créées en

médecine. Si ces critiques sont valides, il s'agit alors d'un troisième moment clé où les valeurs influencent le concept de pathologie.

2.3.1 Le problème de la classe de référence

Kingma (2007) a offert un argument intéressant qui a démontré l'incapacité pour la BST à fournir un concept de santé qui ne ferait pas intervenir les valeurs. Cet argument est basé sur la notion de classe de référence⁴. L'argument consiste à démontrer qu'il est impossible de déterminer une classe de référence appropriée au sein d'une théorie de la santé sans faire intervenir des valeurs. Rappelons que pour la BST, la notion de classe de référence sert à distinguer les groupes d'organismes d'une même espèce sur la base de l'uniformité de leur *design* fonctionnel (Boorse, 1977, p. 562, 1997, p. 7-8). Par exemple, les fonctions reproductives diffèrent en fonction du sexe biologique de l'organisme (p. ex. : un homme qui ne pourrait pas porter d'enfants ne serait pas pour autant dans un état dysfonctionnel). Les fonctions développementales varient également dépendamment de l'âge de l'organisme (p. ex. : la croissance est propre à l'enfance et à l'adolescence). L'âge et le sexe⁵ seraient donc les deux classes de référence qui découpent l'espèce humaine en sous-groupes de *design* fonctionnel

⁴ La classe de référence a également été critiquée par Cooper 2005, Giroux 2009 et Varga 2011, mais sous des angles légèrement différents (voir également Boorse, 2014 pour une réponse aux critiques de Cooper, 2005 et de Giroux, 2009). Je me concentre sur l'argument de Kingma parce que son argument a l'avantage de mettre en évidence l'impossibilité pour Boorse de soutenir que le choix des classes de référence en médecine est indépendant des valeurs.

⁵ Boorse ajoute également que la race pourrait constituer une autre classe de référence en médecine, mais sa position n'est pas catégorique sur le sujet (voir Boorse, 1977, p. 558; 1987, p. 370; 1997, p. 8; et 2014, p. 702).

uniforme. La notion de classe de référence est donc centrale au sein de la BST, étant donné que c'est selon la classe de référence à laquelle l'organisme appartient que l'état d'un organisme peut être dit pathologique ou non. Selon Boorse, ces classes de référence pourraient être identifiées par des méthodes purement empiriques révélant des faits naturels (Boorse, 1977, p. 562, 2014, p. 695). Les différences entre les classes de référence seraient « assez frappantes » (*striking enough*, Boorse, 1977, p. 558) pour légitimer ce découpage de manière empirique.

Or, Kingma défend que la BST doit présupposer la définition des concepts de santé et de pathologie pour être en mesure d'identifier « les bonnes » classes de référence. Cette présupposition que l'âge et le sexe sont les « bonnes » classes de référence reposerait ultimement sur des jugements évaluatifs. En guise d'illustration de la dépendance des classes de référence aux jugements évaluatifs, elle propose d'imaginer une théorie de la santé alternative à la BST qu'elle appelle XST. Dans cette théorie, l'orientation sexuelle serait une autre classe de référence pertinente au sein de l'espèce humaine. Dans la BST, l'homosexualité est considérée comme pathologique, étant donné que les personnes homosexuelles ne réalisent pas leurs fonctions reproductives selon la manière statistiquement typique propre à leur classe de référence (voir Boorse, 1997, p. 11-12⁶). Par contraste, dans la XST, où l'homosexualité est une classe de référence, les personnes homosexuelles sont en santé, étant donné que leurs fonctions reproductives sont réalisées de manière statistiquement typique par rapport à leur classe de référence. Selon Kingma, Boorse ne posséderait pas les ressources pour justifier le choix de la BST par rapport à la XST, étant donné que les faits de la nature ne révèlent pas à eux seuls quels sont les sous-groupes partageant une homogénéité fonctionnelle qui peuvent être découpés au sein d'une espèce. Autrement dit, les faits empiriques ne

⁶ Notons cependant que bien que Boorse conçoive l'homosexualité comme une pathologie, il ne défend jamais l'idée que cette pathologie doive être traitée par la médecine ou la psychiatrie, conformément à la distinction faite entre santé théorique et santé pratique qu'il a développée (voir Boorse, 1975). C'est la signification pratique du concept de santé qui implique ou non un traitement.

nous disent pas si la BST ou la XST est la meilleure théorie, parce qu'il n'est pas possible, seulement par observation, de savoir si l'orientation sexuelle est une classe de référence. Ainsi, selon Kingma, des considérations normatives sont toujours nécessaires pour que ce découpage soit possible.

Dans son article « A Rebuttal on Health » (2014), Boorse a tenté de répondre à cette objection de Kingma en affirmant que son analyse ne choisit pas la BST par contraste avec d'autres alternatives pour des raisons évaluatives. La BST serait plutôt la meilleure description de l'usage médical du concept de santé par les physiologistes. Une théorie telle que la XST, par sa nature hypothétique, ne correspondrait pas à l'usage médical du concept de santé. À partir de ce constat, Boorse conclut : « Thus, I continue to maintain that value-laden choice of a health concept by anyone, *me or medicine*, is a mythical process » (2014, p. 694, les italiques sont les miennes). Cette réponse apparaît cependant insatisfaisante. En effet, il faudrait montrer qu'en plus du fait que la BST analyse l'usage des experts du concept de santé de la manière la plus juste, cet usage correspond aux faits de la nature et ne fait pas intervenir de valeurs (pour une discussion de l'autorité épistémique accordée aux pathologistes, voir Varga, 2015, 2018). Cette seconde exigence est absente du raisonnement de Boorse. En effet, celui-ci démontre seulement que son analyse correspond à l'usage des experts, mais pas que les experts découpent l'espèce humaine en sous-groupes en faisant uniquement référence aux faits empiriques.

La réponse de Boorse et la confusion qu'elle rend manifeste me poussent à reformuler l'argument de Kingma. Ultimement, ce que cet argument met en lumière est le fait que pour identifier la bonne classe de référence, les physiologistes (ou les experts) doivent avoir recours à des jugements de valeur, puisque les faits empiriques seuls ne révèlent pas la manière exacte dont serait découpée l'espèce humaine en sous-groupes de *design* fonctionnel. Comme je vais le montrer dans les sections 2.3.2 et 2.3.3, c'est pour des raisons pratiques ou en fonction de valeurs sociales et culturelles présumées que les

physiologistes circonscrivent certaines classes de référence, et non pas uniquement à partir d'une lecture du monde naturel indépendante des valeurs. Ces deux prochaines sous-sections renforceront donc l'argument de Kingma, qui reposait sur une théorie alternative hypothétique pour contrer la BST. Les cas que je discuterai montreront que même dans la pratique médicale actuelle, des valeurs influencent le découpage des classes de référence. Le problème n'est pas complètement théorique, il a donc des implications pratiques importantes.

2.3.2 Les classes de référence et les pratiques de classification médicale

Dans la section précédente, j'ai mentionné que Boorse, pour se défendre contre la critique de Kingma, avait affirmé que la BST choisissait bien les classes de référence conformément à l'usage du concept fait par les physiologistes. J'ai soutenu que bien qu'il soit possible que Boorse décrive adéquatement le concept de pathologie à l'œuvre dans les théories médicales, il n'est pas certain que les classes de référence utilisées en médecine reflètent un découpage réel de l'espèce humaine. Dans cette section, j'exposerai l'argument de Doust et ses collègues (2017). Ces auteurs montrent comment le choix des classes de référence en médecine implique généralement des considérations normatives quant aux risques de préjudices pour les individus associés à ce choix. Selon ces auteurs, étant donné ce choix, une définition de la pathologie qui ne ferait pas intervenir les valeurs augmenterait le risque de surmédicalisation des problèmes normaux de vie (soit le fait de rendre pathologiques des conditions qui devraient seulement être considéré comme des problèmes normaux de la vie, en dehors du champ médical). Cette mise en garde est fondée sur une analyse des pratiques médicales de catégorisation des classes diagnostiques. Doust et ses collègues écrivent :

An important motivation for the BST was to resist political misuse of concepts of health and disease, such as pernicious forms of medicalization in psychiatry, by insisting that for disease, some identifiable dysfunction must be present. But today, the identification of disease with dysfunction can actually drive new forms of medicalization, which expand the boundaries of disease in ways that are problematic. Focusing on the possibility of naturalistic disease definition becomes problematic where it obscures the ways in which values come into play in real-world practices of defining disease (Doust *et al.*, 2017, p. 352).

Donc, selon elles, même si l'objectif de Boorse était de contrer la médicalisation des problèmes normaux de vie à des fins politiques ou sociales, une définition de la santé qui ne ferait pas intervenir les valeurs pourrait justement servir cette médicalisation.

Selon Doust et ses collègues, cette menace est justifiée par le fait que la BST ne concorde pas avec la manière dont la médecine théorique identifie ce qui constitue une condition pathologique. En effet, la médecine théorique fait appel à des valeurs tout au long du processus de délimitation de la pathologie. En passant en revue trois exemples de pathologies qui ne sont pas controversées, soit le syndrome de Stein-Leventhal (aussi appelé « ovaires polykystiques »), l'insuffisance rénale chronique (dans ce qui suit, IRC) et l'infarctus du myocarde, les auteures montrent que les chercheurs dans le domaine médical doivent avoir recours aux valeurs s'ils veulent établir les limites d'une pathologie particulière (voir aussi Plutynski, 2018, chapitre 2, pour le même genre d'argument avec le cas du cancer). Dans ce qui suit, je vais me concentrer principalement sur l'IRC, étant donné qu'il s'agit de l'exemple le plus parlant d'intrusion des valeurs dans la détermination des pathologies. L'IRC est utilisé par les auteures en guise d'exemple pour illustrer le besoin de recourir aux valeurs pour circonscrire une classe de référence, et ainsi permettre d'identifier une dysfonction. Comme elles le notent, dans sa défense contre ses critiques, Boorse affirme qu'il ne serait pas forcé de justifier davantage les classes de référence de la BST, soit l'âge et le sexe, à moins que des cas réels de disputes médicales existent à propos de

l'identification des classes de référence (Boorse, 2014, p. 695). C'est exactement un cas de ce genre qui est présenté par Doust et ses collègues, qui montrent que les mesures diagnostiques pour identifier un patient atteint de l'IRC dépendent de la classe de référence choisie, et que cette classe de référence est déterminée en partie par des valeurs.

Examinons donc la manière dont l'IRC est catégorisée en médecine. L'IRC est définie comme une réduction du fonctionnement du rein pour une certaine période de temps. Le fonctionnement du rein est mesuré grâce à deux méthodes : le débit de filtration glomérulaire, soit le volume de liquide filtré par le rein par unité de temps, ainsi que la présence d'albumine, une protéine plasmatique, dans l'urine. La mesure « seuil » de ces deux méthodes, qui détermine si une condition est pathologique, est établie en référence au risque de développer des maladies cardiovasculaires et d'aboutir à une insuffisance rénale en phase terminale qui est corrélé à cette mesure. Ces corrélations ont été obtenues en utilisant des études populationnelles de cohorte [*population cohort studies*] qui ont inclus l'âge, le sexe et un grand nombre de groupes ethniques. À la suite de ces études, le seuil officiel établi permettant de distinguer le fonctionnement normal du rein d'un fonctionnement pathologique est de $< 60\text{mL}/\text{min}/1.73\text{m}$. Or, ce seuil impliquait qu'à peu près 50 % des personnes âgées de 75 ans et plus seraient diagnostiquées d'une insuffisance rénale à cause du déclin du fonctionnement rénal associé à l'âge. Cette prévalence est apparue comme trop élevée, ce qui a conduit à l'émergence de critiques dans les écrits médicaux. Ces critiques ont défendu que le risque de développer des maladies cardiovasculaires et l'IRC en phase terminale était pourtant très bas au sein de ce groupe de personnes. Qui plus est, en utilisant le même seuil, des patients plus jeunes, avec la même probabilité de développer des maladies cardiovasculaires et l'insuffisance rénale en phase terminale, ne pourraient pas être diagnostiqués.

Ce qui est en jeu dans cet exemple est la manière dont la classe de référence est déterminée. Le choix d'une classe de référence comme plus inclusive ou plus restreinte aura pour conséquence d'inclure ou d'exclure des patients de la catégorie diagnostique. Par exemple, choisir une classe de référence comme « adultes » par contraste avec plusieurs sous-classes de référence basées sur l'âge (« 18-25 ans », « 25-35 ans », etc.), déterminera quels patients pourront être diagnostiqués ou non en fonction du seuil pathologique impliqué par ce choix. En effet, une mesure particulière sera vue comme le « bon » seuil, qui distinguera entre le niveau de fonctionnement normal du rein du niveau pathologique, en fonction du choix de la classe de référence. Selon Doust et ses collaboratrices, ce ne sont pas seulement les faits de la nature qui disent aux chercheurs quelle est la « bonne » classe de référence. La « bonne » classe de référence est plutôt établie en faisant appel à des critères cliniques *tels que le risque de préjudices pour les patients*.

La manière dont les classes de référence sont déterminées en médecine sert de critique contre la BST. Puisque le risque de préjudice explique en partie le choix d'une classe de référence, la BST est inadéquate pour refléter la manière dont les chercheurs identifient les limites de la pathologie. Doust et ses collègues argumentent donc contre une définition de la santé et de la pathologie indépendante des valeurs. Selon elles, en présentant une définition de la santé qui ne ferait pas intervenir les valeurs, il devient possible de masquer les présuppositions sociales et politiques qui entrent en jeu dans la délimitation du pathologique. L'effacement des valeurs sociales et culturelles présentes dans la définition de la pathologie pourrait augmenter la tendance à la médicalisation, qui était pourtant ce qu'une définition neutre sur le plan des valeurs devait prévenir. En effet, la notion de dysfonction devait servir à limiter l'expansion du concept de pathologie : lorsqu'aucune dysfonction n'était présente, un état ne pouvait être dit pathologique. Ainsi, une définition de la santé devrait viser à reconnaître les valeurs qui entrent en jeu, plutôt que de les « cacher » par l'utilisation de notions neutres sur le plan des valeurs telles que la « dysfonction ». Une définition

de la santé doit donc être transparente quant aux rôles des valeurs dans le processus de catégorisation du normal et du pathologique dans la recherche médicale.

Dans le cas de l'IRC, on voit que les classes de référence sont déterminées en fonction des risques de préjudices encourus pour les patients.es. C'est donc les risques de développer des symptômes qui causent de la souffrance ou des incapacités, voire la mort, qui viennent motiver le découpage de la population humaine en sous-groupes. Ce n'est donc pas une pure lecture empirique du monde qui explique les classes de référence de la médecine, comme le soutient Boorse.

2.3.3 Les classes de référence de la BST et les critiques du dualisme homme-femme

Pour illustrer davantage en quoi le choix des classes de référence peut être influencé par les valeurs, je vais m'inspirer des récents travaux qui portent sur le dualisme sexuel afin de montrer que même une classe de référence qui apparaît « assez frappante », telle que le sexe biologique, ne peut être identifiée sans le recours à des jugements de valeur.

Blackless et coll. (2000), Callahan (2009), Fausto-Sterling (2000) et Amoretti et Vassallo (2017) prennent pour cible l'approche dualiste des sexes et des genres acceptée en médecine et plus récemment dans la « gender specific medicine » (voir spécifiquement Amoretti et Vassallo, 2017 pour une analyse de ce courant émergent en médecine). L'approche dualiste se comprend comme la croyance selon laquelle il existe une démarcation claire entre des catégories binaires de sexes et de genres. La médecine reconnaît donc deux sexes biologiques, les sexes masculin et féminin, qui

sont généralement associés à deux genres binaires, les genres masculin et féminin. Selon ces auteurs, ce dualisme strict, qu'il concerne le genre ou le sexe, serait un présupposé accepté implicitement ou explicitement dans les recherches en médecine. Or ce présupposé serait partiellement infondé. Comme l'écrit Blackless : « If, however, one relinquishes an a priori belief in complete genital dimorphism, one can examine sexual development with an eye toward variability rather than bimodality » (Blackless *et al.*, 2000, p. 151). Ces auteurs se basent sur les récents travaux ayant porté sur les personnes intersexes pour remettre en question le dimorphisme sexuel. Dans ce qui suit, je vais me concentrer uniquement sur les critiques du dimorphisme sexuel qui concernent le sexe biologique, étant donné que c'est cet aspect qui est lié à la classe de référence du sexe présentée par Boorse dans la BST.

La distinction entre les sexes biologiques masculin et féminin repose généralement sur le fait de posséder des petits gamètes, les spermatozoïdes, associés au sexe masculin, ou de plus gros gamètes, les ovules, associés au sexe féminin. Cependant, il existe plusieurs traits sexuels qu'il convient de considérer d'un point de vue biologique : le sexe génétique (46 chromosomes XX contre 46 chromosomes XY), le sexe gonadique (les testicules contre les ovaires), le sexe gamétal (les spermatozoïdes contre les ovules), le sexe génital (les organes génitaux internes et externes associés au sexe gamétal), le sexe hormonal (testostérone contre oestrogène) et le sexe somatique (caractéristiques sexuelles secondaires telles que la pilosité ou la distribution de gras). Ainsi, l'homme biologique typique possède des chromosomes XY et des testicules localisés dans le scrotum. Les testicules produisent les spermatozoïdes qui peuvent être transportés via le canal déférent à l'urètre et être éjaculés. La taille du pénis varie de 2,5 à 4,5 cm à la naissance. Durant le développement fœtal, les testicules produisent l'hormone de régression müllérienne, la testostérone et la dihydrotestosterone, alors que l'activité testiculaire juvénile permet la puberté masculinisante. Quant à elle, la femme biologique typique possède des chromosomes XX, des ovaires qui permettent une puberté féminisante et un oviducte qui permet l'acheminement des ovocytes

produits par les ovaires vers l'utérus. Elle possède également un col de l'utérus et un canal vaginal, des lèvres vaginales internes et externes et un clitoris dont la taille varie de 0,20 à 0,85 cm. (Blackless *et al.*, 2000, p. 152). Bien que chez la majorité des êtres humains, l'ensemble de ces « niveaux » de sexe biologique concorde, les personnes intersexes, qui représentent de 1,7 à 2 % de la population, manifestent une combinaison des différents sexes biologiques masculins et féminins (Amoretti et Vassalo, 2018; Callahan, 2009; Fausto-Sterling, 2000).

Un exemple de personne intersexe est Maria Patiño, une athlète féminine, qui s'est vu refuser le droit de participer à des compétitions sportives dans la catégorie « femme » parce qu'il a été découvert que, bien qu'elle possédait des organes génitaux féminins et s'identifiait comme une femme, elle possédait des chromosomes XY, soit les chromosomes sexuels masculins (Fausto-Sterling, 2000). Maria Patiño représente un cas d'intersexualité parmi d'autres. En effet, il existe plusieurs types d'intersexualité, causés par différents facteurs de développement (pour un tableau qui représente les types les plus connus d'intersexualité, voir Fausto-Sterling, 2000, p. 52). L'un des types d'intersexualité les plus connus est l'hyperplasie congénitale des surrénales (HCS), causée par une variation génétique associée à l'enzyme impliquée dans la production des hormones stéroïdes, l'enzyme 21-hydroxylase. La variation cause une sous-production de cortisol et d'aldostérone. Chez les filles affectées par l'HCS, leurs chromosomes sont XX et elles naissent avec des organes reproductifs féminins qui sont la plupart du temps fertiles, mais avec des organes génitaux externes masculins. Certaines formes de HCS peuvent entraîner des complications, comme une perturbation de la métabolisation du sel, potentiellement létal si elles ne sont pas traitées à la cortisone. Un autre cas est le syndrome de Turner, soit des filles qui possèdent seulement un chromosome X (XO). Le syndrome de Turner implique que les filles ne posséderont pas d'ovaires, posséderont des organes génitaux féminins, mais ne manifesteront pas de caractéristiques sexuelles secondaires (Callahan, année, p. 69 et 181; Fausto-Sterling, 2000, p. 73-76). Le syndrome de Klinefelter, quant à lui,

survient chez les garçons. Il est caractérisé par un caryotype 47 XXY et entraîne la stérilité. Le mosaïsme, quant à lui, désigne la condition d'une personne qui possède plus d'une population de cellules, c'est-à-dire que le contenu génétique des cellules varie d'une cellule à l'autre. Par exemple, une personne pourrait avoir des cellules dont les chromosomes sont composés de 45 XO et de 46 XY (voir Sandberg *et al.*, 2017). L'HCS, le syndrome de Turner, le syndrome de Klinefelter et le mosaïsme ne sont que des exemples de la variabilité retrouvée chez les personnes intersexes.

Dans les exemples précédents, la combinaison d'un sexe génital féminin et d'un sexe génétique masculin démontre que « les faits de la nature » ne permettent pas de trancher en faveur de deux catégories clairement distinctes de sexe biologique. Le fait de découper l'espèce humaine en deux sous-groupes distincts, les classes de référence « hommes » et « femmes », implique le recours à des choix en plus d'une observation des faits. Les implications de ces travaux qui remettent en question le dimorphisme sexuel strict peuvent s'exporter au concept de classe de référence présent dans la BST. Dans le cadre de la théorie de Boorse, les personnes intersexes seraient classées comme pathologiques. En effet, si l'on accepte la classe de référence du sexe traditionnel, qui découpe l'espèce humaine en deux sous-groupes de sexe biologique différent, « femme » et « homme », alors toute personne qui ne développe pas les caractéristiques sexuelles typiquement associées à l'une de ces deux classes de référence sera désignée comme étant dans une condition pathologique, parce que ses « parties » (gènes, cellules, organes, etc.) ne fonctionnent pas de manière statistiquement typique selon une efficacité statistiquement typique. La totalité des personnes intersexes se retrouvera donc dans la catégorie « pathologique » étant donné que leur développement sexuel est différent du développement sexuel statistiquement typique de leur classe de référence. Par exemple, une personne présentant l'HCS, qui possède des chromosomes XX (chromosomes féminins) et des organes reproducteurs féminins (ovaires et utérus), sera dite pathologique étant donné que ses organes génitaux externes sont masculins. La présence d'organes génitaux externes masculins est une déviation par rapport à la

manière statistiquement typique dont l'appareil reproducteur féminin contribue à la reproduction, ce qui rend l'HCS pathologique dans le cadre de la BST.

Un premier point qu'il convient de noter est que la BST semble être une bonne description de la manière dont la médecine découpe l'espèce humaine en sous-groupes en fonction du sexe biologique. En effet, en médecine, les personnes intersexes appartiennent à la catégorie des « troubles du développement sexuel » [*disorders of sex development* » ou « *disorder of sex differentiation* »] (voir Amoretti et Vassallo, 2017). La médecine conçoit ainsi la condition intersexe comme une déviation pathologique par rapport au développement sexuel normal qui se comprend selon une différenciation sexuelle binaire. C'est dire que la classe de référence qu'est le sexe biologique, divisée selon les catégories « homme » et « femme », joue un rôle réel de classification en médecine et détermine ce qui est considéré comme normal ou pathologique. Ainsi, du point de vue de la reconstruction rationnelle opérée par Boorse (voir le chapitre 1), la classe de référence du sexe biologique de la BST représente de manière juste la théorie médicale. Or, l'enjeu réside plutôt du côté des raisons qui motivent ce découpage dans la théorie médicale elle-même. En effet, il convient de savoir si cette classe de référence correspond à la « découpe réelle » du monde naturel ou si des valeurs interviennent dans cette découpe. Autrement dit, est-ce que seules des observations empiriques permettent d'établir que l'espèce humaine se divise en deux catégories biologiques, « homme » et « femme »? Face à cette question, la position de Boorse quant au fait que les valeurs n'influencent pas le choix des classes de référence est intenable. En effet, comme le mettent bien en évidence Amoretti et Vassallo (2017), rien n'indique que le fait d'être intersexe devrait en lui-même impliquer un état pathologique (au sens où cette condition n'entraîne pas nécessairement de préjudice pour l'individu en termes de symptômes incapacitants ou de longévité). Bien qu'il soit vrai que certaines formes de développement intersexe soient associées à des pathologies, telles que le syndrome de Klinefelter qui entraîne la stérilité et le HCS qui peut être associé à des perturbations métaboliques, la condition elle-même n'entraîne

pas nécessairement de tels effets. Un individu affecté par le HCS pourrait ne jamais souffrir de perturbations métaboliques et être fertile. Qui plus est, les individus présentant les caryotypes 47 XXY et 47XYY ne sont généralement pas diagnostiqués parce qu'ils ne présentent aucun symptôme préjudiciable de cette divergence génétique (Blackless, 2000).

Le même raisonnement élaboré par Kingma concernant l'homosexualité s'applique ici, mais pour des classes de référence réellement utilisées en médecine. Il semble en effet qu'il faille accepter d'abord que la classe de référence représentant un dualisme sexuel est la « bonne » classe de référence pour que les personnes intersexes soient exclues du domaine du normal. Or, considérer qu'il s'agit de la bonne classe de référence ne peut se faire sans le recours à certaines présuppositions telles que le dimorphisme sexuel. Le raisonnement est alors circulaire, fait intervenir des valeurs, et ne repose pas seulement sur des observations empiriques. À la lumière de cette remise en question de la validité des catégories sexuelles binaires, il est juste de conclure que le choix des classes de référence en médecine est informé par des valeurs sociales et culturelles en plus d'être basées sur des observations empiriques, étant donné que le « bon » découpage de l'espèce humaine en sous-groupes partageant une uniformité fonctionnelle n'est pas révélé par des données seules sur le développement sexuel (au chapitre 4, je discute à nouveau cet exemple et en tire des implications quant à l'idéal de la science indépendante des valeurs).

Boorse admet déjà que ce qui est délimité sur un continuum relève de conventions, comme c'est le cas de la performance d'une fonction dont la variation est continue (voir la section 2.2). Si certaines des classes de référence se retrouvent sur un continuum, comme il semble que ce soit le cas avec le sexe biologique, alors Boorse doit admettre que la délimitation des classes de référence relève également de conventions et fait donc intervenir les valeurs. L'exemple des personnes intersexes montre que la catégorie « sexe » est plutôt un continuum. Créer des catégories à partir de ce

continuum implique des choix, motivés en partie par des valeurs. La définition de la BST, parce qu'elle implique le recours à la notion de classe de référence, n'est donc pas indépendante des valeurs sociales et culturelles.

Conclusion : la présence de valeurs au sein de la BST

Dans ce chapitre, je me suis intéressée à la manière dont la BST se comprenait comme une théorie « objective », « neutre » ou « indépendante des valeurs ». Chez Boorse, cette affirmation est soutenue par au moins deux arguments : 1) l'utilisation de méthodes scientifiques pour distinguer le normal du pathologique et 2) la scientificité du concept de pathologie. Après avoir clarifié cette thèse, j'ai montré que Boorse reconnaissait pourtant l'influence de valeurs à au moins deux moments dans sa théorie : 1) le choix des buts de l'organisme et 2) la démarcation entre un fonctionnement bas et un fonctionnement dysfonctionnel. Cette reconnaissance va à l'encontre de sa thèse selon laquelle les concepts de santé et de pathologie sont indépendants des valeurs. J'ai par la suite discuté l'influence des valeurs au sein de la BST avec la notion centrale de classe de référence. Dans un premier temps, j'ai présenté, avec le cas hypothétique de la XST développé par Kingma, que la BST avait du mal à justifier ses classes de références sans recourir à des jugements évaluatifs. J'ai ensuite utilisé deux exemples concrets, l'IRC et le dualisme homme-femme, pour illustrer la manière dont les valeurs pouvaient intervenir dans le processus de catégorisation des classes de référence en médecine. Comme je l'ai mis en lumière dans le chapitre 1, les buts de l'organisme, le seuil de fonctionnement pathologique et la classe de référence sont trois notions clés de la BST. J'ai montré dans ce chapitre que ces trois notions étaient influencées par les

valeurs sociales et culturelles. Il me semble donc difficile d'accepter la thèse de Boorse qui soutient que la BST est totalement indépendante des valeurs.

Références du chapitre 2

- Amoretti, M. C. et Vassallo, N. (2015). Against sex and gender dualism in gender-specific medicine. Dans U. Mäki, I. Votsis, S. Ruphy et G. Schurz (dir.), *Recent developments in the philosophy of science : EPSA13 Helsinki* (p. 357-367). s.l. (Suisse) : Springer, Cham.
- Amoretti, M. C. et Vassallo, N. (2017). Introduction : Philosophy of sex and gender in gender medicine. *Topoi*, 36(3), 473-477.
- Ananth, M. (2008). *In defense of an evolutionary concept of health : Nature, norms, and human biology*. Burlington, VT : Ashgate Publishing Group.
- Blackless, M. (2000). Recommendations from interACT : Advocates for intersex youth regarding the List of issues for the United States for the 59th session of the Committee against torture. *Human Biology*, 12, 151-166.
- Blackless, M., Charuvastra, A., Derryck, A., Fausto- Sterling, A., Lauzanne, K. et Lee, E. (2000). How sexually dimorphic are we? Review and synthesis. *American Journal of Human Biology*, 12(2), 151-166.
- Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & public affairs*, 5(1), 49-68.
- Boorse, C. (1976a). What a theory of mental health should be. *Journal for the Theory of Social Behaviour*, 6(1), 61-84.
- Boorse, C. (1976b). Wright on functions. *The Philosophical Review*, 85(1), 70-86.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44(4), 542-573.

- Boorse, C. (1987). Concepts of health. Dans D. Van de Veer et T. Regan (dir.), *Health care ethics : An introduction* (p. 359-393). Philadelphie : Temple University Press.
- Boorse, C. (1997). A rebuttal on health. Dans J. M. Humber et R. F. Almeder (dir.), *What is disease?* (p. 1-134). Totowa, NJ : Humana Press.
- Boorse, C. (2002). A rebuttal on functions. Dans A. R. Ariew, R. Cummins et M. Perlman (dir.), *Functions : New essays in the philosophy of psychology and biology* (p. 63-113). Oxford : Oxford University Press.
- Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, 39(6), 683-724.
- Callahan, G. (2009). *Between XX and XY : Intersexuality and the myth of two sexes*. Chicago : Chicago Review Press.
- Doust, J., Walker, M. et Rogers, W. A. (2017). Current dilemmas in defining the boundaries of disease. *Journal of Medicine and Philosophy*, 42(4), 350-366.
- Fausto-Sterling, A. (2000). *Sexing the body : Gender politics and the construction of sexuality*. New York : Basic Books.
- Kingma, E. (2007). What is it to be healthy? *Analysis*, 67(2), 128-133.
- Kingma, E. (2010). Paracetamol, poison, and polio : Why Boorse's account of function fails to distinguish health and disease. *The British Journal for the Philosophy of Science*, 61(2), 241-264.
- Kingma, E. (2014). Naturalism about health and disease : Adding nuance for progress. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 39(6), 590-608.
- Murphy, D. (2006). *Psychiatry in the scientific image*. Cambridge, MA : MIT Press.
- Plutynski, A. (2018). *Explaining cancer : Finding order in disorder*. Oxford : Oxford University Press.

- Rogers, W. A. et Walker, M. J. (2017). The line-drawing problem in disease definition. *Journal of Medicine and Philosophy*, 42(4), 405-423.
- Sandberg, D. E., Pasterski, V. et Callens, N. (2017). Introduction to the special section : Disorders of sex development. *Journal of Pediatric Psychology*, 42(5), 487-495.
- Schwartz, P. H. (2007). Defining dysfunction : Natural selection, design, and drawing a line. *Philosophy of Science*, 74(3), 364-385.
- Varga, S. (2011). Defining mental disorder : Exploring the “natural function” approach. *Philosophy, Ethics, and Humanities in Medicine*, 6(1), 1-10.
- Varga, S. (2018). Epistemic authority, philosophical explication, and the bio-statistical theory of disease. *Erkenntnis*, 85, 937-956.

CHAPITRE 3

LES APPROCHES ALTERNATIVES

Introduction

Dans les chapitres précédents, j'ai présenté la BST de Boorse et ses problèmes. À la suite de mon analyse, il est apparu que la BST n'était pas satisfaisante pour définir le trouble mental. Cette insatisfaction pourrait nous pousser à rechercher des approches alternatives au naturalisme de Boorse. Deux autres approches ont été développées dans les écrits : le constructivisme et l'approche hybride. Le constructivisme s'oppose au naturalisme en ce qu'il pose les jugements évaluatifs comme critères pour distinguer le normal du pathologique. L'approche hybride se veut une tentative de conciliation entre le naturalisme et le constructivisme, en ce qu'elle présente le trouble mental comme à la fois défini par des valeurs et par des faits scientifiques. Dans ce chapitre, j'exposerai à grands traits ces deux approches, pour ensuite en montrer leurs limites respectives.

3.1 Se tourner vers le constructivisme?

3.1.1 Présentation de l'approche constructiviste

Face à l'échec de l'approche naturaliste de Boorse, il peut sembler prometteur d'adopter plutôt l'approche à laquelle on l'oppose dans les écrits, à savoir le constructivisme/normativisme. Selon le constructivisme, il serait impossible de définir la santé et la pathologie de manière indépendante des valeurs sociales et culturelles (voir Boorse, 1975, p. 50). À l'image du terme « naturalisme », le terme constructivisme peut revêtir plusieurs significations. Bien qu'il existe plusieurs théories constructivistes du trouble mental, il est possible de brosser à grands traits le portrait de cette conception. La caractéristique principale qui lie ces différentes approches est le fait de poser que le critère pour départager le normal du pathologique se trouve sur le plan des jugements de valeur, et non des faits empiriques (Lemoine, 2017, p. 35). De manière générale, le constructivisme comprend la santé et la pathologie en référence à des normes personnelles ou sociales. Cela ne veut pas dire que les constructivistes nient qu'il existe une base physiologique ou psychologique associée aux troubles mentaux. Ils défendent plutôt qu'il n'existe pas de science du fonctionnement normal de l'être humain qui soit indépendante des normes ou des jugements de valeur. Murphy résume la conception constructiviste et son opposition avec la conception naturaliste ainsi :

[...] constructivists assert that to call a condition a disease is to make a judgment that someone in that condition is undergoing a specific kind of harm that we explain in terms of bodily processes. But the bodily processes are not objectively malfunctioning; rather, they are merely judged by us to be unusual or abnormal because they depart from some shared, usually culturally specific,

conception of human nature. The crucial difference between the positions then is that for naturalists, diseases are objectively malfunctioning biological processes that cause harms. For constructivists, diseases are harms that we blame on some biological process because it causes the harm, not because it is objectively dysfunctional (Murphy, 2020).

La distinction principale entre le normativisme et le naturalisme est donc la croyance selon laquelle une notion indépendante des valeurs viendrait expliquer la distinction entre le normal et le pathologique. Un exemple peut aider à illustrer la position constructiviste : l'obésité (Murphy, 2020). Le constructivisme admettrait qu'il soit possible de découvrir des faits scientifiques à propos de l'obésité tels que des relations statistiques entre l'obésité et la pression sanguine ou l'espérance de vie. La découverte de ces faits scientifiques est rendue possible parce que nous avons d'abord identifié l'obésité comme un problème. Dans cette optique, l'obésité n'est jamais « découverte » comme une pathologie. Ce sont plutôt les faits scientifiques qui lui sont associés qui sont découverts. La présupposition selon laquelle l'obésité est une déviation permet de lui apposer son étiquette de « pathologie ». Cette étiquette ne vient donc pas de l'observation de la nature, mais d'un jugement de valeur par rapport à ce que devrait être le bon fonctionnement du corps humain. Le même type d'exemple peut être donné dans le domaine de la psychiatrie avec la dépression majeure. Nous pouvons recueillir des informations empiriques sur les phénomènes neuropsychologiques qui sous-tendent les symptômes de la dépression, comme l'humeur dépressive, une variation dans le poids et des troubles du sommeil, mais cela ne revient pas à dire que la dépression est identifiée comme une pathologie pour des raisons empiriques. Pour les constructivistes, certaines conditions sont considérées comme des pathologies pour des raisons qui relèvent du domaine des valeurs.

Les approches constructivistes varient selon le type de normes impliqué dans le concept de trouble mental (Murphy, 2020). Dans ce qui suit, je vais reprendre la catégorisation des différentes formes de constructivisme présentée par Lemoine (2017), qui découpe

trois manières distinctes de concevoir les normes impliquées dans les concepts de santé et de pathologie. Bien que ces différentes formes de constructivisme soient beaucoup plus riches et complexes que ce que je vais présenter, je vais me concentrer principalement sur la manière dont les normes sont intégrées au sein de ces différentes approches. Cette accentuation me permettra de bien contraster le constructivisme avec la position hybride (que je vais présenter et critiquer dans la section 3.3) et le naturalisme de Boorse discuté dans les chapitres 1 et 2.

Selon Lemoine, trois types de constructivisme cohabitent dans les écrits en philosophie de la médecine et de la psychiatrie : le constructivisme axiologique, le constructivisme pragmatique et le constructivisme eudémoniste (Lemoine, 2017, p. 34-38). Avant de détailler chacune de ces approches, il est important de mentionner qu'une autre forme de constructivisme existe : le « constructivisme fort », comme développé, entre autres, par Szasz (voir Agich, 1997 pour cette appellation). Puisque j'ai déjà développé cette version du constructivisme au chapitre 1 dans le cadre du dialogue entretenu entre Boorse et Szasz, je ne m'y attarderai pas ici. Rappelons simplement que pour Szasz, le trouble mental est un mythe, qui dénote uniquement la déviance sociale et morale, ce qui fait de lui un tenant de l'approche constructiviste en psychiatrie¹. Cela étant dit, je vais présenter les trois formes de constructivisme pour ensuite en présenter les critiques dans la section 3.1.2.

Le constructivisme axiologique comprend la pathologie comme ce qui produit des états qui gênent ou qui sont désagréables. Le concept de pathologie sert alors à décrire et à

¹ Une nuance doit cependant être apportée ici : Szasz peut en effet être dit constructiviste fort quant au concept de santé mentale. Or sa conception de la santé somatique est objectiviste, en ce qu'il considère qu'une lésion physiologique est le critère qui permet de distinguer le normal du pathologique, critère qui se veut objectif. C'est parce que le trouble mental n'est pas associé à une lésion physiologique que Szasz soutient l'inexistence « objective » du trouble mental. En l'absence de lésion physiologique, le trouble mental ne peut être que l'expression d'une déviance vis-à-vis les normes sociales (voir Murphy, 2006, p. 25 pour cette nuance).

expliquer un *pattern* de phénomènes qui sont associés à ces expériences négatives, mais aussi à motiver l'action en réponse à ces phénomènes. Étant donné que ces conditions sont jugées désagréables, on cherche à s'en débarrasser. Dans ce cadre, la médecine cherche à obtenir des connaissances sur les causes de ces états dans une visée thérapeutique. Selon Engelhardt, représentant de l'approche axiologique, plusieurs significations différentes seraient associées au concept de pathologie, selon le contexte dans lequel le terme est utilisé : légal, moral, thérapeutique ou expérientiel. Le concept de pathologie impliquerait des devoirs et des responsabilités de la part du personnel médical, un rôle particulier pour le patient affecté, etc., en fonction du contexte social au sein duquel le terme est utilisé (Engelhardt, 1975, p. 126-127). Selon cette vision, des pathologies telles que le cancer, la tuberculose ou la schizophrénie existent, mais en tant que *pattern* d'explication utilisé dans des contextes sociaux particuliers (Engelhardt, 1975, p. 136). Par exemple, le fait de considérer la schizophrénie comme une pathologie implique toujours un jugement de valeur quant aux états désirables auxquels devraient correspondre les individus. La schizophrénie est jugée comme une pathologie parce que ses symptômes sont perçus comme indésirables. Considérer la schizophrénie comme une pathologie implique alors des actions du point de vue de l'intervention médicale, du rôle du patient et de la prise en charge des professionnels de la santé.

Engelhardt est également connu pour ses travaux sur la masturbation, conçue comme une pathologie au 18^e et 19^e siècle (Engelhardt, 1974). À cette époque, la masturbation était perçue comme causant, entre autres, la dyspepsie (un ensemble de troubles digestifs), l'épilepsie, la cécité, la perte de mémoire, des maux de tête et une diminution générale de la santé (Engelhardt, 1974, p. 236). La masturbation était donc comprise comme une pathologie au regard de ses effets indésirables. Par ailleurs, le concept d'« onanisme » était utilisé pour expliquer différents signes et symptômes interreliés et corrélés à un seul mécanisme causal responsable de la pratique répétée de la masturbation et de ses effets indésirables (Engelhardt, 1974, p. 237). Selon Engelhardt,

dans le cas de la masturbation, un comportement sexuel jugé moralement condamnable, était perçu comme entraînant des dysfonctions sur le plan physique. À partir de ce qui était compris comme l'idéal type biologique (compris en référence à des normes), un certain type de sexualité (génitale et à l'intérieur d'un couple), les états qui s'en écartaient étaient jugés comme pathologiques, et des descriptions causales expliquant comment ils apparaissaient étaient développées dans une visée thérapeutique.

Une conception similaire qui s'ancre également dans l'approche du constructivisme axiologique a été développée plus récemment par Rachel Cooper (2002, 2005; voir Reznick, 1987 pour une approche similaire). Pour Cooper:

By "disease" we aim to pick out a variety of conditions that through being painful, disfiguring, or disabling are of interest to us as people. This class of conditions is by its nature anthropocentric and corresponds to no natural class of conditions in the world (Cooper, 2005, p. 22).

Pour Cooper, trois éléments doivent être respectés pour qu'une condition ou un état soit jugé pathologique : 1) une condition doit être jugée comme mauvaise par la personne affectée; 2) la personne affectée doit être considérée comme malchanceuse, c'est-à-dire qu'il y a de bonnes raisons d'espérer de ne pas être atteint par ce trouble; 3) un traitement existant ou potentiellement existant doit pouvoir être envisagé (Cooper, 2005, p. 23). Pour Cooper, la combinaison de ces trois éléments doit être nécessaire et suffisante pour que l'on puisse parler de pathologie. La première condition sert à distinguer les réelles pathologies des simples variations au sein de la population humaine (p. ex. : les cheveux roux, les yeux verts, les génies). Qui plus est, Cooper insiste sur le fait que cette condition doit être préjudiciable pour l'individu, et non pas pour la société, ou pour la majorité des individus affectés par cette condition (bien qu'elle admette que de déterminer si un individu est affecté par sa condition peut s'avérer être une tâche complexe [Cooper, 2005, p. 26-28 pour une discussion]). La deuxième condition, celle de la malchance, sert à distinguer les pathologies des

conditions qui sont déplaisantes ou souffrantes, mais normales (p. ex. : l'accouchement). Pour que le fait d'avoir une condition soit considéré comme malchanceux, il faut que le fait d'espérer ne pas avoir eu la condition soit raisonnable. Cooper donne l'exemple de la perte de mobilité chez les personnes âgées. Il est raisonnable de s'attendre à ce qu'un homme de 90 ans ne soit plus capable de marcher comme lorsqu'il avait 30 ans. Or, il n'est pas raisonnable de s'attendre à ce qu'un homme de 30 ans soit incapable de marcher, ce pour quoi l'on peut qualifier sa condition de « malchance » (pour une discussion, voir Cooper, 2005, p. 29-32). La troisième condition, la potentialité de traitement, sert à distinguer les pathologies de simples malchances (p. ex. : la pauvreté, la laideur, etc.). Cette condition n'implique pas qu'un traitement existe à l'heure actuelle pour soigner une condition donnée, mais uniquement qu'il soit raisonnable de penser qu'un traitement puisse être développé. Cela ne veut pas dire non plus qu'une personne affectée par un trouble doive nécessairement chercher à être traitée si ce traitement est disponible : une personne pourrait refuser un traitement (p. ex. : quelqu'un qui refuse les soins médicaux pour traiter son cancer a quand même une maladie, parce que le traitement est disponible). Selon Cooper, pour que cette troisième condition soit respectée, il faut donc uniquement que la potentialité de ce traitement soit présente. Cooper mentionne également que la condition doit potentiellement être traitable d'un point de vue social. Par exemple, il est possible d'imaginer que les comportements violents soient traitables, mais on ne les considère pas comme des pathologies (Cooper, 2005, p. 34). Cooper et Engelhardt sont des représentants de l'approche du constructivisme axiologique parce que leur conception fait référence à deux composantes : d'abord, le caractère désagréable ou préjudiciable du trouble, et ensuite une référence à la réponse thérapeutique entraînée par la présence du trouble. Des différences entre les deux approches demeurent : pour Engelhardt, le préjudice est davantage jugé par la société, tandis que pour Cooper, il doit être jugé par l'individu. Qui plus est, pour Engelhardt, la réponse thérapeutique doit généralement être mise en place, tandis que pour Cooper, le caractère potentiel d'une réponse thérapeutique suffit.

Le constructivisme pragmatique, pour sa part, comprend la pathologie comme une défaillance ou un échec dans l'action d'un individu, généralement associés à un risque de souffrance accru. Cette approche a été défendue principalement par Clouser *et al.* (1981), Whitbeck (1981) et Agich (1983, 1997, 2002) en philosophie de la médecine, de même que Fulford (1989) en philosophie de la psychiatrie. Pour Agich, représentant exemplaire de l'approche pragmatique, cette conception met l'accent sur la nature pratique du concept de pathologie, c'est-à-dire sur le fait que le concept accomplit diverses fonctions dans le langage médical. L'approche pragmatique vise donc à mettre en lumière les différentes fonctions que joue le concept, les normes et les valeurs impliquées dans ces usages, puis la critique de cet usage (Agich, 1997, p. 238). Dans cette optique, le concept de pathologie désigne l'expérience d'incapacité à accomplir ses actions intentionnelles de la vie courante ressentie par un individu. Pour Agich, les actions intentionnelles de la vie courante sont définies par l'individu en fonction de ses valeurs et de ses croyances (Agich, 1997, p. 236). Par exemple, si un individu souffre d'une dépression majeure, il ne parvient généralement pas à accomplir les actions de la vie ordinaire telles que se rendre au travail, maintenir une vie sociale, etc. Si ce sont des tâches valorisées par l'individu, la dépression s'accompagne donc d'un échec dans l'action. Ce concept implique également de la part du corps médical une tentative d'amélioration de la condition de l'individu au regard de sa capacité d'action (Agich, 1997, p. 226). Par exemple, la prise en charge thérapeutique visera à redonner à l'individu la capacité d'action qu'il valorise. Le constructivisme pragmatique insiste donc sur le caractère incapacitant d'une condition par rapport aux intentions d'action des individus et sur la réponse du corps médical face à cette incapacité.

Le constructivisme eudémoniste comprend quant à lui la santé comme la capacité à réaliser ses buts vitaux. L'auteur le plus connu ayant défendu cette approche est Lennart Nordenfelt, avec sa théorie holistique de la santé (Nordenfelt, 1987, 1995, 2001, 2007; voir Giroux, 2010, p. 111-144 pour une présentation plus détaillée et une discussion;

voir aussi Megone, 1998, 2000 pour une approche similaire à Nordenfelt). Nordenfelt s'écarte de la conception boorséenne de la santé et de son approche « part-functionalism » en concevant la santé dans une perspective globale où seul l'individu peut être dit en santé ou malade (par contraste avec ses comportements ou ses organes). Pour Nordenfelt, « A is completely healthy if, and only if, A has the ability, given standard circumstances, to reach all his or her vital goals. » (Nordenfelt 2007, p. 7). Je vais expliquer les principaux concepts de cette approche. D'abord, l'individu, compris comme un sujet d'action, est au centre de la théorie holistique de Nordenfelt. La capacité [*ability*] réfère à une possibilité d'action pour l'individu si les opportunités requises à la réalisation de l'action lui sont fournies. Or, ce ne sont pas toutes les opportunités requises qui sont pertinentes pour penser la santé; seules les circonstances standard le sont. Les circonstances standard dont fait mention Nordenfelt dans la définition ne concernent pas des normes statistiques sur ce qui constituerait un environnement normal, mais sont déterminées en fonction de normes culturelles et sociales qui forment le contexte de vie de l'individu. Pour qu'un individu soit en santé, il doit pouvoir réaliser ses buts vitaux dans un ensemble de circonstances jugé standard par la société (p. ex. : il n'est pas nécessaire de pouvoir respirer sur la lune pour être en santé, puisque cette circonstance n'est pas jugée comme standard par la société). Qui plus est, Nordenfelt précise la notion de capacité en distinguant les capacités de premier ordre et les capacités de second ordre. La capacité de premier ordre désigne le fait de pouvoir, dans le moment présent, réaliser une action. La capacité de second ordre désigne une capacité potentielle, c'est-à-dire que la capacité d'acquérir, par formation ou entraînement, une capacité de premier ordre (p. ex. : je pourrais ne pas savoir lire présentement [capacité de premier ordre], mais être capable d'apprendre à lire suite à un enseignement [capacité de second ordre]). Pour Nordenfelt, ce sont les capacités de second ordre qui importent pour penser la santé (1995, p. 50-53).

Quant aux buts vitaux, ils désignent les buts considérés essentiels par l'individu, c'est-à-dire que lorsque les buts vitaux sont atteints, l'individu se trouve dans un état de

bonheur minimal et durable (ces buts peuvent être différents selon les individus, mais tendent toutefois à être assez partagés). Nordenfelt met donc l'accent sur l'expérience de la santé et de la maladie dans un contexte social donné. Cette expérience peut correspondre à un ressenti d'accomplissement ou de frustration par rapport à la capacité (de second ordre) d'un individu à agir dans une certaine direction. Si l'individu ressent une frustration dans sa capacité à réaliser ses buts vitaux, il pourra être considéré comme souffrant d'une pathologie. Ainsi, pour Nordenfelt, un individu peut être dit en santé s'il a la capacité de second ordre de réaliser les actions qui contribuent à son bonheur minimal et durable dans un environnement standard. À l'inverse, il sera dit dans un état pathologique s'il ne possède pas cette capacité. Notons donc que Nordenfelt accorde une priorité à l'expérience subjective d'un problème (souffrance ou handicap) par rapport à la description des états psychologiques ou biologiques associés à ce problème. Autrement dit, si une condition n'est pas liée à un préjudice ressenti par l'individu, elle ne devrait pas être considérée comme une pathologie. Nordenfelt accorde ainsi la priorité à l'expérience vécue de la santé et de la pathologie par rapport à la description d'une dysfonction « organique », ce qui fait de sa conception de la santé une théorie holistique et normative.

Le constructivisme est donc une approche qui a revêtu plusieurs formes dans les écrits. Cette conception de la santé et de la pathologie apparaît attrayante pour deux raisons principales. D'abord, les critiques formulées par les tenants de l'approche constructiviste envers l'approche objectiviste ont bien permis d'en montrer les limites, principalement en ce qui concerne ses ambitions de fournir une définition indépendante des valeurs et des intérêts humains. D'ailleurs, la plupart des critiques qui ont visé la conception naturaliste de Boorse ont été formulées par des auteurs appartenant au courant constructiviste (les auteurs rejettent généralement l'idée d'un concept de santé théorique et indépendant des valeurs). Deuxièmement, sous un angle plus positif, le constructivisme fournit des outils pour reconnaître le rôle joué par les valeurs sociales et culturelles dans le concept de trouble mental. En effet, le constructivisme s'intéresse

généralement aux processus sociaux qui mènent à la catégorisation, et à la manière dont cette catégorisation change à travers le temps et selon les sociétés et cultures. Cette approche est donc intéressante parce que l'accent est mis sur les pratiques sociales qui mènent à la catégorisation d'un état comme pathologique. Cet intérêt a généralement été dissocié de l'objectivisme, qui vise plutôt à fournir un ou des critères objectifs permettant de rendre compte de cette catégorisation à partir de la méthode de l'analyse conceptuelle. Or le constructivisme n'est pas sans problème, ce que je montrerai dans la section suivante.

3.1.2 Présentation des problèmes associés au constructivisme

Trois problèmes principaux sont associés au constructivisme (voir par exemple Wakefield, 1992b; certains de ces problèmes sont également repris par Murphy, 2006 et Ereshefsky, 2009). D'abord, selon Wakefield, le fait que des troubles mentaux soient associés à des jugements négatifs ne signifie pas nécessairement que les troubles mentaux ne soient que le reflet de jugements négatifs. Autrement dit, montrer que des valeurs sociales et culturelles sont impliquées dans la classification des troubles mentaux (ou des pathologies somatiques) n'entraîne pas nécessairement qu'il n'y aurait pas aussi une notion empirique et objective nécessaire à l'attribution d'une condition pathologique à un patient. Wakefield défend que l'approche constructiviste n'a pas fait la démonstration qu'il n'existait aucun concept scientifique et objectif rattaché au concept général de trouble mental, même si elle démontre bien la présence de valeurs lui étant généralement associées. Deuxièmement, il existe plusieurs conditions que la société dévalorise qui ne sont pas pour autant considérées comme des troubles mentaux. En effet, la petitesse, la timidité, la mauvaise fortune et la pauvreté ne sont

pas considérées comme des troubles mentaux, même si ce sont des états dévalorisés par la société. Il semblerait donc qu'une distinction existe entre ces deux types de cas. Selon Wakefield, l'approche constructiviste est incapable de rendre compte de nos intuitions quant à la distinction entre des états dévalorisés qui sont pathologiques et des états dévalorisés qui ne le sont pas. Autrement dit, si les normes sociales sont ce qui explique la différence entre la pathologie et la santé, comment se fait-il que des conditions soient dévalorisées, mais pas considérées comme des pathologies (p. ex. : le cancer par contraste avec la pauvreté)? L'approche constructiviste resterait sans réponse face à cette question. La troisième critique soulevée par Wakefield concerne l'incapacité des constructivistes à exclure du domaine du pathologique certains phénomènes tels que la dissidence politique en Union soviétique ou la drapétomanie, exemples hautement problématiques pour cette approche (Wakefield, 1992b, p. 377). Ainsi, selon Wakefield, puisque les constructivistes défendent que la pathologie doit être comprise en référence aux valeurs, si une société dévalorise certaines conditions, les constructivistes ne seraient pas en mesure d'affirmer qu'il s'agit d'un mauvais jugement ou d'une erreur. Si l'on reprend l'exemple de la drapétomanie dans une société raciste, les constructivistes ne pourraient pas dire que la drapétomanie n'est pas un trouble mental, indépendamment du contexte culturel et social. Il serait uniquement possible de dire que maintenant, dans nos sociétés occidentales contemporaines, la drapétomanie n'est pas pathologique (voir aussi Glackin, 2019 pour cet exemple). Pour Wakefield, sans concept scientifique indépendant des valeurs, les constructivistes ne peuvent pas affirmer que la dissidence politique ou la drapétomanie ne sont pas des troubles mentaux puisque ces conditions étaient jugées comme pathologiques par la société.

Les trois arguments de Wakefield s'accompagnent de l'idée selon laquelle le trouble mental impliquerait nécessairement une notion indépendante des valeurs, la notion de dysfonction. Comme nous le verrons dans la section suivante, pour Wakefield, la réponse aux limites de l'approche constructiviste est l'ajout d'une notion de

dysfonction scientifique. Or pour le moment, je vais discuter ces trois critiques mises de l'avant par Wakefield pour elles-mêmes. La première critique cible l'idée selon laquelle ce n'est pas tout de démontrer que des jugements de valeur influencent notre concept de maladie; il faut également démontrer qu'un concept indépendant des valeurs comme celui de dysfonction *n'influence pas* notre concept, ou ne devrait pas influencer notre concept. Or contre cette critique de Wakefield, il me semble que l'approche constructiviste a grandement participé à démontrer que l'existence d'un concept de dysfonction indépendant des valeurs était extrêmement difficile à démontrer. Dans le chapitre précédent, j'ai présenté les critiques qui avaient visé la notion de dysfonction chez Boorse. Dans la suite de ce chapitre, je présente les critiques concernant cette notion chez Wakefield. En ce sens, la première critique de Wakefield me semble viser un homme de paille. En effet, les constructivistes participent généralement à déconstruire l'idée d'une composante parfaitement neutre qui définirait la santé et la pathologie. Il serait néanmoins possible d'argumenter que même si Boorse et Wakefield ont échoué à élaborer un concept de dysfonction indépendant des valeurs qui soit satisfaisant, cette quête ne doit pas être abandonnée. En effet, l'une des intuitions objectivistes est que seul un concept indépendant des valeurs peut assurer une distinction objective entre le normal et le pathologique. Pour répondre à ce possible contre-argument, dans le chapitre suivant, je fournis un argument inspiré du débat sur les valeurs en science pour justifier le rejet de la quête d'un concept de dysfonction indépendant des valeurs.

La deuxième critique de Wakefield me semble à nouveau imprécise. Pour Wakefield, le fait que les approches constructivistes fondent la distinction entre le normal et le pathologique dans des jugements évaluatifs implique qu'elles ne peuvent pas distinguer les conditions dévalorisées qui sont des pathologies des conditions dévalorisées qui sont seulement considérées comme des problèmes normaux de vie. Cette critique me semble reposer sur un présupposé erroné : la pathologie n'est pas qu'une condition dévalorisée par la société. Comme je l'ai présenté dans la section précédente, les

approches constructivistes sont beaucoup plus subtiles et élaborées que ce que suppose cette critique. Par exemple, l'approche du constructivisme axiologique défendue par Cooper ajoute la potentialité d'un traitement et la malchance au critère de dévalorisation. Qui plus est, bien qu'il soit vrai de dire que la pathologie implique une dévalorisation, Cooper situe ce jugement chez l'individu. C'est l'individu qui doit percevoir sa condition négativement pour qu'elle puisse être considérée comme une pathologie. Il est donc faux de dire que l'ensemble des approches constructivistes sont incapables de rendre compte de la distinction entre les jugements négatifs sur des conditions qui sont des pathologies et des conditions qui sont simplement socialement dévalorisées. Il est toutefois vrai de dire que pour la plupart des approches constructivistes, la conception de la santé et de la maladie implique un certain relativisme par rapport à une société donnée (p. ex. : Nordenfelt et les circonstances considérées comme « standard » par la société). Toutefois, cette relativité s'accompagne généralement d'autres critères qui visent à départager ces cas problématiques.

La troisième critique de Wakefield me semble plus fondamentale. Pour Wakefield, il est difficile pour les constructivistes de condamner des conceptions de la santé et de la maladie qui ont émergé dans des contextes historiquement ou socialement différents, comme la drapétomanie et l'homosexualité. Même si les constructivistes fournissent généralement de bonnes descriptions du rôle des valeurs dans la définition de la santé mentale, cette approche ne fournit pas ou peu de critères *prescriptifs* qui viendraient délimiter les concepts de santé et de trouble mental. Par exemple, on peut expliquer dans le cadre constructiviste comment la drapétomanie et l'homosexualité ont pu être conceptualisées comme des pathologies dans des sociétés racistes ou homophobes. Il est plus difficile, à même le cadre constructiviste, de dire que ces jugements étaient mauvais ou erronés. Plus généralement, bien que le constructivisme me semble juste dans sa reconnaissance du rôle joué par les jugements évaluatifs lorsque des conditions sont classées comme normales ou pathologiques, cette approche ne fournit pas ou peu

de ressources pour répondre aux défis soulevés par le mouvement antipsychiatrique et l'existence de plusieurs catégories diagnostiques psychiatriques controversées (voir aussi Powell et Scarffe, 2019a pour une critique similaire). Autrement dit, bien qu'il soit clair que des valeurs influencent les concepts de santé et de trouble mental grâce aux travaux des constructivistes, cette approche ne vise généralement pas à *justifier* quelles valeurs devraient influencer le concept de trouble mental. En effet, le constructivisme est avant tout une approche critique ou descriptive, qui vient mettre en lumière la présence de valeurs qui influencent le concept de trouble mental. Les définitions offertes par le constructivisme sont généralement situées historiquement et socialement, et ne visent pas à justifier, au-delà de ce contexte historique et social ou des intuitions communément partagées, pourquoi des définitions passées (p. ex. : une définition qui inclut l'homosexualité comme une pathologie) sont mauvaises par rapport à celles proposées, outre par la description de nos intuitions. Or, la question de la justification des valeurs en jeu n'est pas de nature descriptive, ni purement historique ou sociologique. Il s'agit d'une question normative, voire éthique, portant sur les valeurs qui *devraient ou ne devraient pas* influencer notre concept de trouble mental. À cet effet, il n'est pas clair que le constructivisme donne des outils pour faire face à l'utilisation abusive des catégories psychiatriques sur certaines populations (p. ex. : dissidents politiques, groupes socialement opprimés, etc.). Ce n'est pas tout d'affirmer que des valeurs sont en jeu dans la signification du concept de trouble mental : il faut pouvoir défendre leur rôle légitime et montrer à quel moment et selon quels critères cette influence devient illégitime (p. ex. : il faudrait pouvoir dire que l'homophobie n'est pas une valeur acceptable et qu'une définition du trouble mental ne devrait pas être influencée par des valeurs homophobes). Cette étape est généralement manquante au sein des approches constructivistes, ce qui me paraît une limite importante. C'est avec la notion de dysfonction que Wakefield pense pouvoir dépasser cette limite importante de l'approche constructiviste. Comme je le montrerai dans la section suivante, il n'est pas certain que cette stratégie soit la bonne. Dans les chapitres suivants, j'élaborerai une conception de l'objectivité qui a pour but de réconcilier la

reconnaissance des valeurs en jeu dans le concept de trouble mental et le besoin de juger certaines conceptions de la santé et de la pathologique comme erronées.

3.2 Se tourner vers la théorie de la dysfonction préjudiciable?

À la vue des problèmes présentés précédemment, d'aucuns pourraient être tentés d'adopter, plutôt qu'une théorie purement naturaliste (comme développée par Boorse), ou purement constructiviste (comme développée par Engelhardt, Agich et Nordenfelt, entre autres), une théorie mixte ou hybride qui intègre à la fois des normes biologiques et sociales. Une telle théorie a été élaborée par Jerome Wakefield à partir des années 1990 en philosophie de la psychiatrie. Wakefield propose de dépasser l'opposition entre naturalistes et normativistes en définissant le trouble mental comme une dysfonction préjudiciable [*harmful dysfunction*] (DP, dans le reste du texte). Selon Wakefield, le concept de trouble mental fait référence à la fois à une *norme physiologique*, soit le *design* de l'espèce sélectionné par la sélection naturelle, et à une *norme sociale et culturelle*, soit le préjudice dont le sujet ou ses proches font l'expérience. Ces deux normes seraient nécessaires pour identifier un état comme pathologique. Wakefield écrit :

I argue that disorder lies on the boundary between the given natural world and the constructed social world; a disorder exists when the failure of a person's internal mechanisms to perform their functions as designed by nature impinges harmfully on the person's well-being as defined by social values and meanings. The order that is disturbed when one has a disorder is thus simultaneously biological and social; neither alone is sufficient to justify the label disorder. (Wakefield, 1992b, p. 373)

Cette tentative d'intégration des valeurs sociales et culturelles et des faits scientifiques au sein de la définition de la santé mentale a valu à Wakefield une notoriété importante au sein des écrits en philosophie de la psychiatrie.

Avec Boorse, Wakefield s'inscrit dans le courant de l'approche objectiviste des fonctions naturelles (*natural function objectivism*, Varga, 2015), qui défend que la composante centrale du trouble mental est une dysfonction de certains mécanismes qui composent l'esprit humain. Cette dysfonction peut être identifiée de manière objective par une investigation scientifique indépendante des intérêts humains. Deux éléments principaux différencient cependant l'approche de Wakefield et celle de Boorse. D'abord, pour Wakefield, il n'existe pas deux concepts distincts de maladie, le concept théorique et le concept pratique, mais un seul. Ce concept unique combine une composante axiologique, le préjudice, et une composante descriptive, la dysfonction. La théorie de la fonction adoptée par Wakefield est également à distinguer de celle adoptée par Boorse. Wakefield défend une théorie sélectionniste de la fonction alors que Boorse adopte la théorie « orientée vers des buts » de la fonction (voir le chapitre 1). Dans ce qui suit, je vais exposer la théorie de la DP en mettant l'accent sur quatre éléments principaux : les notions de fonction naturelle et de dysfonction (section 3.2.1), l'indépendance des valeurs de la notion de dysfonction (section 3.2.2), la thèse de la boîte noire essentialiste (section 3.2.3) et la notion de préjudice (section 3.2.4). Dans la section 3.3, j'identifierai les limites de la conception de Wakefield.

3.2.1 La notion de dysfonction et la théorie de la sélection naturelle

Dans le cadre de la DP, la notion de dysfonction se comprend en référence à la notion de fonction naturelle, renvoyant au *design évolutionniste* de l'espèce humaine. Dans ce qui suit, je vais présenter la théorie de la fonction adoptée par Wakefield, et comment la notion de dysfonction y est comprise. Wakefield adopte ce qui est appelé, dans les écrits sur les fonctions, l'approche étiologique ou de « l'effet sélectionné », développée par Millikan (1989) et Neander (1991). Dans cette approche, la fonction d'un trait est ce pour quoi il a été sélectionné au cours de l'évolution. Le *design évolutionniste* de l'espèce humaine est donc l'ensemble des fonctions sélectionnées au cours de son histoire évolutive. La sélection naturelle filtre graduellement les traits et comportements au sein d'une population et tend à conserver ceux qui sont avantageux au regard d'un environnement donné, grâce au succès reproductif (ou *fitness*) des individus possédant les traits et comportements avantageux (Buss *et al.*, 1999, p. 535). Par exemple, avoir peur des serpents a pu être avantageux dans un environnement passé sauvage et hostile. Les individus étant effrayés par les serpents ont pu, en moyenne, survivre et se reproduire davantage que leurs confrères et consœurs plus téméraires. Ainsi, la peur des serpents a pu être sélectionnée. Les traits et comportements qui sont « sélectionnés » sont appelés des « adaptations ». En se basant sur ce processus de sélection naturelle, Wakefield défend qu'une fonction naturelle d'un mécanisme s'explique par les effets passés de celui-ci sur la survie et la reproduction des individus possédant ces mécanismes. L'action passée de la sélection naturelle explique alors la présence actuelle des mécanismes (Wakefield, 2001, p. 348). Ainsi, selon Wakefield, c'est le processus de la sélection naturelle qui explique l'existence et le maintien de certains mécanismes typiques au sein de l'espèce humaine, soit les fonctions naturelles de l'espèce humaine, ou le *design évolutif* (Wakefield, 1992b, p. 383, 1999a). Par exemple, selon cette approche, la fonction du cœur est de pomper le sang, car le fait de

pomper le sang est ce qui explique la présence et le maintien actuel du cœur au sein de l'espèce humaine. C'est en raison de cet effet bénéfique que cet organe a été sélectionné, c'est-à-dire que le fait de pomper le sang a conféré un avantage en termes de survie et de reproduction aux porteurs de ce trait. Cet avantage évolutif explique la structure du cœur et les mécanismes qui assurent que le sang soit pompé. Pomper le sang est donc la fonction naturelle du cœur pour Wakefield (et ceux qui adoptent cette position).

Pour Wakefield, c'est grâce au cadre de la sélection naturelle que l'on peut départager des traits ou comportements qui sont valorisés au plan personnel ou social, mais qui n'ont pas de fonctions naturelles. Par exemple, le nez humain permet de maintenir en place des lunettes, ce qui peut être avantageux. Cependant, ce n'est pas sa fonction naturelle, étant donné que la présence du nez chez l'humain n'est pas explicable par le fait qu'il permette de maintenir des lunettes en place. Le nez n'a pas été sélectionné parce qu'il permettait le port de lunettes : ce n'est donc pas sa fonction naturelle. Il en va de même pour le cœur. Le son que produit le battement du cœur peut être utile dans le cadre d'exams médicaux, mais le bruit du cœur n'est pas explicable en termes d'avantages évolutifs passés. Dans l'histoire évolutive passée, le bruit produit par le battement du cœur n'a pas été sélectionné. Le fait de faire du bruit est donc un effet corrélé [*by-product*] et non pas la fonction naturelle du cœur (Wakefield, 1999a, p. 471).

Dans le cadre de la théorie de la fonction comme effet sélectionné, une dysfonction est le résultat d'un mécanisme qui contrevient aux fonctions naturelles pour lequel il a été sélectionné (Wakefield, 1992b, p. 383). En guise d'illustration, un cœur qui ne pompe pas le sang de manière suffisante serait dans un état dysfonctionnel étant donné que cet état s'éloigne de la fonction naturelle sélectionnée du cœur. Pour revenir à l'exemple du nez, un nez avec une forme qui ne permettrait pas de soutenir des lunettes ne serait pas dysfonctionnel, parce que ce n'est pas la fonction naturelle du nez que de soutenir

des lunettes (Wakefield, 1992a, p. 381). De la même manière, un cœur dont le bruit produit par le battement serait différent de la normale ne serait pas dans un état dysfonctionnel, parce qu'émettre un battement sonore n'est pas la fonction sélectionnée du cœur. Même si nous préférons que le nez soutienne les lunettes ou que le cœur émette un bruit particulier, ne pas se conformer à ces préférences n'implique pas une dysfonction. En effet, ce ne sont pas les fonctions naturelles du nez ou du cœur : ils n'ont pas été sélectionnés pour remplir ces fonctions. C'est cette notion de dysfonction désormais clarifiée qui joue un rôle central dans la définition du trouble mental. Wakefield définit en effet le trouble mental comme une dysfonction préjudiciable, où la dysfonction se comprend dans le cadre de la théorie de la fonction sélectionnée.

3.2.2 L'indépendance des valeurs de la notion de dysfonction

Bien que Wakefield intègre à sa théorie une composante évaluative avec la notion de préjudice, la manière dont est comprise *l'objectivité (ou la neutralité) de la dysfonction* dans ce contexte est relativement similaire à Boorse (à la différence que la théorie de la fonction adoptée par Wakefield est sélectionniste). Pour Wakefield, ce concept de dysfonction est un concept « purement scientifique », « factuel », que l'on peut découvrir empiriquement (Wakefield, 1992a, p. 236, 1992b, p. 373). Wakefield écrit à cet effet : « A disorder is different from a failure to function in a socially preferred manner precisely because a dysfunction exists only when an organ cannot perform as it is naturally (i.e., independently of human intentions) supposed to perform » (Wakefield, 1992b, p. 381). Selon Wakefield, parce que le concept de dysfonction est ancré dans la théorie de l'évolution, une théorie bien établie en science, c'est un concept purement factuel et scientifique qui réfère à quelque chose de « réel » (Wakefield,

1999b, p. 472; voir aussi Wakefield, 2003, p. 974, 978). Qui plus est, la théorie de l'évolution par sélection naturelle a l'avantage d'être un processus non intentionnel, qui ne dépend pas des intérêts humains. La conception de Wakefield de la neutralité du concept de fonction est donc similaire à celle de Boorse, en ce qu'elle prend pour point de départ l'ancrage du concept de fonction au sein d'une théorie scientifique, ici la théorie de l'évolution. Parce que le concept de fonction est un concept scientifique et que la théorie dans lequel il s'inscrit est non intentionnelle, ce concept est indépendant des valeurs sociales et culturelles. Bien sûr, dire que quelqu'un a une condition dysfonctionnelle implique une évaluation normative de cette condition (p. ex. : le mécanisme M ne fonctionne pas comme il le devrait). Or, cette évaluation se rapporte à la norme contenue dans le *design* fonctionnel sélectionné par l'évolution passée. Bien que l'évaluation d'une condition comme fonctionnelle ou dysfonctionnelle comporte une part d'évaluation la norme de référence est physiologique, et donc indépendante des normes sociales et culturelles (Wakefield, 2001, p. 352 et 353)².

Pour Wakefield, trois arguments principaux donnent du poids à l'idée selon laquelle les valeurs ne peuvent rendre compte de l'entièreté du concept de trouble mental. D'abord, comme suggéré dans la section 3.1.2, l'ancrage de la notion de trouble mental à la notion de dysfonction est essentiel pour pouvoir expliquer la distinction entre des conditions dévalorisées, mais non pathologiques, et des conditions que nous considérons réellement comme pathologiques. En appui à cette idée, Wakefield écrit :

Values cannot be the whole story regarding the meaning of 'disorder' because both professionals and laypersons distinguish negatively valued behaviors that are disorders from those that are not. For example, illiteracy, although disvalued and harmful in our society, is not in itself considered a disorder, whereas a

² Il est à noter que Wakefield ne nie pas que les valeurs sociales et culturelles puissent influencer nos attributions fonctionnelles (voir Wakefield, 2003). Or, cela n'implique pas que le concept de fonction en lui-même soit constitué de valeurs sociales et culturelles.

similar condition believed to be due to lack of ability to learn to read because of some internal neurological or psychological dysfunction is considered a disorder. Grief is seen as normal whereas similarly intense sadness states not triggered by real loss are seen as disorders. An analysis of “disorder” must explain such distinctions among negatively evaluated behavior. They are not explained by a pure value account (Wakefield, 2003, p. 972).

C'est donc l'existence de conditions dévalorisées, mais qui ne sont pas considérées comme des troubles mentaux, qui donne du poids à la notion de dysfonction. Wakefield affirme que la dysfonction permet de faire la différence entre ces deux types de cas, ce qu'une théorie purement constructiviste ne pourrait pas faire. La dysfonction permet de rendre compte du fait que nous distinguons généralement ces deux types de cas.

Dans un deuxième temps, Wakefield observe que dans certains cas, on considère des dysfonctions comme neutres ou positives. La dysfonction et les valeurs personnelles ou sociales et culturelles seraient donc à distinguer :

Further evidence for a value-free “dysfunction” component is that we understand how to apply “dysfunction” independently to value-neutral conditions. Numerous ‘benign anomalies’ of physical structure and genetic coding clearly are dysfunctions but are evaluatively neutral, thus not classified as disorders. For example, simple angiomas (blood vessels that reach the skin) are considered dysfunctions in which blood vessel growth has gone astray, and a substantial percentage of most men’s sperm cells are malformed, yet both these conditions are harmless, thus not classified as disorders. Indeed, we apply the notion of “dysfunction” even to evaluatively positive conditions that are not disorders, such as the immune-cell receptor dysfunction that is harmless because of redundancy but apparently protects against AIDS (Wakefield, 2003, p. 273).

L'existence de certaines dysfonctions neutres, voire de dysfonctions valorisées, plaide en faveur de la distinction entre « dysfonction » et « valeur ». Une dysfonction peut donc être valorisée, dévalorisée ou neutre. L'idée selon laquelle le trouble mental serait

donc uniquement de l'ordre des valeurs ne peut pas rendre compte de ces différentes formes de jugement par rapport à la dysfonction. Dysfonction et valeurs doivent donc être séparées.

Pour finir, l'existence de variations culturelles dans les modes de vie suggère une différence entre la dysfonction et les valeurs, puisque nous ne jugeons pas des formes de vie différentes de notre culture comme pathologiques, même si nous les dévalorisons parfois :

Moreover, we often correct our views of disorder based on cross-cultural evidence that allows us to form hypotheses about functions and dysfunctions that transcend and may even go against our values. For example, our culture does not value polygamy, but we consider it not to be a failure of natural functioning, thus not disordered, based partly on cross-cultural evidence (Wakefield, 2003, p. 273).

Encore une fois, il semble que nous faisons la différence entre dysfonction et conditions socialement dévalorisées. Pour ces trois raisons, Wakefield pense qu'un concept purement évaluatif ne peut pas rendre compte de nos intuitions par rapport au trouble mental. C'est donc la notion de dysfonction qui viendrait limiter l'inclusion des conditions seulement dévalorisées au sein du domaine du pathologique : si ces conditions ne sont pas accompagnées d'une dysfonction, elles ne peuvent pas être dites pathologiques. Rappelons que pour Wakefield, les fonctions naturelles ont été sélectionnées par l'évolution, un processus indépendant des intérêts humains. Les notions de fonction et de dysfonction, puisqu'elles s'expliquent par ce processus, sont donc elles aussi indépendantes des intérêts humains. Il y a ainsi une composante qui ne relève pas des intérêts humains dans la définition de trouble mental, la dysfonction, et cette condition permet de venir départager des cas controversés.

Pour mieux comprendre les arguments de Wakefield, il convient de noter que la majeure partie de ces arguments reposent sur la méthode de l'analyse conceptuelle, et ne sont donc pas puisés à même une théorie scientifique (Aucouturier et Demazeux, 2013; Lemoine, 2013). À cet effet, Wakefield écrit :

The HD analysis attempts to explain these classificatory intuitions. The use of natural functions rather than social functions or individually intended functions is dictated *by the data of our intuitions about disorder*, not by a value preference (Wakefield, 2003, p. 975, les italiques sont les miennes; voir aussi p. 971).

Ce sont donc les intuitions du sens commun qui fournissent le matériau initial pour définir le trouble mental. Pour Wakefield, l'intuition principale qui sous-tend les attributions de troubles mentaux est l'idée selon laquelle quelque chose ne fonctionne pas bien chez l'individu (Wakefield, 1992a, p. 240). La notion de dysfonction vise donc d'abord à rendre compte de cette intuition et est identifiée par l'analyse conceptuelle. Wakefield décrit la méthode de l'analyse conceptuelle ainsi : « In a conceptual analysis, proposed accounts of a concept are tested against relatively uncontroversial and widely shared judgments about what does and does not fall under the concept » (Wakefield, 1992a, p. 233). L'analyse de la DP est donc dans un premier temps descriptive plutôt prescriptive ou stipulative (voir Lemoine, 2021 pour une discussion plus détaillée). Elle vise à rendre compte de nos intuitions sur le trouble mental. Rappelons que cette méthode est très similaire à celle employée par Boorse (voir le chapitre 1), à la différence que Wakefield s'intéresse aux intuitions sur le sens clinique et pratique du trouble mental (contrairement à Boorse, qui s'intéresse uniquement au sens théorique, voir Demazeux, 2015, p. 77). Qui plus est, Wakefield analyse à la fois les intuitions des experts et des non-experts, alors que dans une optique naturaliste, Boorse considère uniquement les intuitions des experts comme pertinentes pour fonder la définition de la pathologie. Qui plus est, Boorse s'intéresse seulement à une classe d'experts particuliers, les physiologistes, alors que Wakefield s'intéresse aux intuitions d'un

ensemble d'experts (les psychiatres, les travailleurs sociaux, les physiologistes, etc.) en plus de s'intéresser aux intuitions du sens commun.

Une précision doit être faite à ce stade pour comprendre comment la théorie de la fonction comme effet sélectionné trouve sa place dans cette conception. Ce n'est pas l'entièreté de l'analyse de Wakefield qui fait usage de l'analyse conceptuelle. À cet effet, Lemoine distingue trois étapes qui mènent à la DP :

- 1) Le fait de dire que le trouble mental implique la notion de dysfonction;
- 2) Le fait d'affirmer que les notions de fonction et de dysfonction sont des termes scientifiques et causaux;
- 3) Le fait de soutenir que la meilleure conception de la fonction est ancrée dans la théorie de l'évolution (Lemoine, 2021).

Comme le souligne Lemoine, seulement les étapes 1 et 2 font appel à l'analyse conceptuelle, alors que l'étape 3 se rapproche davantage d'une explication. L'une des manières de critiquer la théorie de Wakefield a été de trouver des contre-exemples à son analyse, c'est-à-dire des exemples de troubles mentaux qui n'impliqueraient pas de dysfonction ou qui n'impliqueraient pas de préjudice (Bolton, 2007, p. 122). Pour être conformes à la méthode de Wakefield, ces critiques doivent seulement concerner les étapes 1 et 2. L'étape 3, par contraste, vise davantage à expliquer les notions de fonction et de dysfonction dans un cadre scientifique existant et se distancie par là de la méthode de l'analyse conceptuelle. L'étape 3) est donc plus prescriptive, en ce que Wakefield défend que la théorie de l'évolution est le meilleur cadre pour penser le concept de fonction en médecine et en psychiatrie. La manière de critiquer Wakefield sur ce point sera donc de défendre que la théorie de la fonction sélectionnée n'est pas la meilleure approche pour penser le trouble mental, ce que je ferai dans la section 3.3.

3.2.3 La DP, l'essentialisme et la boîte noire

En plus de la théorie de la fonction sélectionnée et de l'objectivité de la notion de dysfonction, un autre élément important de la DP est « l'essentialisme de la boîte noire » (*black box essentialism*, voir Wakefield 1997, 1999a, 2000, 2004, 2011). Par cette idée, Wakefield vient spécifier le caractère scientifique et indépendant des valeurs de la notion de dysfonction. Selon lui, les concepts de fonction et de dysfonction sont ultimement des concepts scientifiques essentialistes, qui doivent être compris sur le modèle des espèces naturelles (*natural kind concept*) développé par Putnam (1975) et Searle (1983). De manière générale, l'essentialisme de la boîte noire peut se comprendre ainsi :

In sum, in a black-box-essentialist definition, the surface properties of a base set are used to pick out a possibly unknown underlying essence, and the essence is used as the necessary and sufficient criterion for category membership of new instances, whether or not the new instances share the surface properties of the base set. (Wakefield, 1997, p. 658).

Cette conception est essentialiste parce qu'elle postule que l'appartenance à une catégorie (p. ex. : celle de jade, chrome, lapin, ou encore fonction) est déterminée par le fait de partager une hypothétique essence commune (Wakefield, 2004, p. 79). Lorsqu'appliquée au trouble mental, cette idée implique que la dysfonction, composante nécessaire du concept, puisse être définie à partir de cas paradigmatiques non controversés (dépression, schizophrénie, etc.) qui partageraient la/les mêmes causes sous-jacentes ou essences. Des manifestations différentes en termes de symptômes ou de conditions ne sont alors pas incompatibles avec le fait de partager la même essence. Wakefield offre l'analogie de l'eau (H₂O) pour illustrer cette idée. L'eau peut se manifester selon plusieurs formes (solide, liquide, gazeuse, dans un

mélange avec d'autres types de liquides, etc.). Or, toutes ces manifestations ne changent pas le fait qu'il s'agit d'eau puisqu'elles sont expliquées par la même structure moléculaire : H₂O. H₂O est donc l'essence de l'eau, en ce qu'elle explique et sous-tend les différentes caractéristiques superficielles de l'eau. Selon Wakefield, il pourrait en aller de même pour le trouble mental, défini en partie par la notion de dysfonction. Toutes les manifestations de fonction sont explicables en référence à l'action de la sélection naturelle. La notion de dysfonction se comprend donc en référence à cette essence commune. En ce sens, tous les cas de dysfonction réfèrent à une essence commune, la sélection naturelle passée (Wakefield, 2000, p. 36). Bien que certaines variations dans la manifestation du trouble mental soient observables, cela n'empêche pas qu'une essence commune soit partagée par toutes ces manifestations : l'explication par l'action de la sélection naturelle passée. La différence entre l'eau et les fonctions naturelles est que l'eau est définie en termes de substance (sa composition chimique), alors que les fonctions naturelles sont définies en termes d'explications des effets (c'est-à-dire par les effets de certains mécanismes sur la survie et la reproduction passés des organismes, qui expliquent leur maintien) (Wakefield, 2000, p. 37). L'explication de la dysfonction est l'action passée de la sélection naturelle.

Par ailleurs, pour Wakefield, pour qu'un concept scientifique relève de la boîte noire essentialiste, l'essence d'un concept n'a pas besoin d'être connue, mais seulement postulée. Pour mieux comprendre cette idée, revenons à la méthode employée par Wakefield, l'analyse conceptuelle. Cette méthode révèle une intuition communément partagée à propos du trouble mental, celle d'un dysfonctionnement. La méthode scientifique, quant à elle, permet de révéler les causes de ce dysfonctionnement. La sélection naturelle explique les fonctions normales et nous permet de comprendre les dysfonctions comme les mécanismes qui s'écartent du fonctionnement normal sélectionné. La découverte scientifique explique alors l'intuition communément partagée (voir Demazeux, 2015, p. 81 pour cette clarification). Se référant toujours à l'analogie de l'eau, Wakefield explique qu'il n'était pas nécessaire de connaître la

composition chimique essentielle de l'eau pour savoir que les formes gazeuse et liquide relèvent de la même substance. Avec l'idée de la boîte noire, cette même composition peut être seulement postulée ou inférée, plutôt que connue avec certitude (Wakefield, 2000, p. 36). À cet effet, Wakefield écrit :

The proposed concepts are *black box* because, rather than defining concepts by specific theoretical properties (e.g., “Water is H₂O”), such concepts postulate a theoretical explanatory structure and remain agnostic on its identity, which may be unknown (e.g., “Water is anything that has the same substance – essence as the clear thirst-quenching liquid in the lakes and rivers”) (Wakefield, 2004, p. 79).

Ceux qui utilisent ce type de concept peuvent alors rester agnostiques quant à la nature exacte de l'essence, qui pourra être ultérieurement découverte par la science. Wakefield se sert de cette idée pour soutenir qu'à travers le temps, le concept de trouble mental est resté le même. Même si des théoriciens ont réfléchi sur le trouble mental avant la découverte de la théorie de l'évolution, il s'agissait du même concept parce l'essence était la même. La seule limite à leur réflexion était l'ignorance de son explication scientifique, la théorie de la sélection naturelle. Qui plus est, Wakefield reconnaît l'état embryonnaire de nos connaissances scientifiques sur le fonctionnement normal de l'esprit humain, tel qu'il a été sélectionné. Selon Wakefield, dans l'état actuel des connaissances sur l'évolution de l'esprit humain, l'identification des dysfonctions se fait à partir de jugements intuitifs basés sur des « éléments de preuve circonstanciels » (Wakefield, 1999a, 2003). Wakefield précise peu ce que signifient des « éléments de preuve circonstanciels ». Selon moi, cela signifie que bien que les hypothèses évolutionnistes sur les traits et mécanismes sélectionnés ne soient pas définitives jusqu'à maintenant, cela n'empêche pas de considérer ces hypothèses comme raisonnables et de les accepter jusqu'à ce qu'on ait une meilleure connaissance de l'histoire évolutive humaine :

Obviously, in our present state of ignorance, we judge that there is a failure of some internal mechanism from circumstantial evidence that makes such a hypothesis overall the most plausible, just as, long before we understood etiology, we correctly recognized blindness and paralysis as disorders based on the circumstantial evidence of an individual's inability to see or move under conditions in which it was presumed that normally functioning individuals would be able to see or move, and where alternative explanations seemed unlikely (Wakefield, 2003, p. 971; voir aussi Wakefield, 1992b, p. 383, 2001, p. 356).

Ainsi, pour Wakefield, il n'est pas nécessaire de connaître avec certitude les fonctions naturelles sélectionnées de l'esprit humain pour pouvoir inférer qu'une dysfonction est présente dans le cas d'un diagnostic de trouble mental, ou à l'inverse, qu'une dysfonction n'est pas présente, et que le diagnostic est erroné. Pour Wakefield, cette ignorance n'empêche pas de considérer la dysfonction comme un concept scientifique et indépendant des valeurs, ni d'affirmer que la meilleure théorie de la fonction pour rendre compte de nos intuitions sur le trouble mental est celle de l'effet sélectionné.

3.2.4 La DP et la notion de préjudice

Bien que les travaux de Wakefield aient mis l'accent sur la notion de dysfonction, il est important de rappeler que selon la DP, deux composantes sont nécessaires pour que l'analyse du trouble mental soit complète : la dysfonction *et* le préjudice. Pour Wakefield, la notion de dysfonction, bien que nécessaire pour départager les cas controversés en psychiatrie et en médecine, n'est pas suffisante. Par exemple, un donneur de rein, qui accepterait de donner son rein à quelqu'un en ayant besoin, serait dans une condition dysfonctionnelle parce que les êtres humains ont probablement été

sélectionnés pour avoir deux reins. Or conformément à nos intuitions, parce que cette personne ne subit pas de préjudice dû à cette dysfonction, elle n'aurait pas à proprement parler un trouble (Wakefield, 1992b, p. 384). Ainsi, pour Wakefield, ce ne sont pas toutes les dysfonctions qui sont considérées comme des troubles ou des pathologies. La notion de préjudice vient alors compléter la signification du trouble mental en ajoutant un autre critère nécessaire relevant des valeurs. Autrement dit, le préjudice est ce qui rend compte du caractère évaluatif du concept de trouble mental. Cette notion est à comprendre en termes de conséquences négatives sur l'individu, évaluées au regard des standards socioculturels d'une société (Wakefield, 1992b, p. 374). Dans cette optique, un individu dyslexique placé dans une société lettrée sera considéré comme présentant un trouble mental (tenant pour acquis que la dyslexie est une dysfonction), alors que le même individu placé dans une société illettrée sera considéré comme seulement dysfonctionnel (et donc sans trouble mental). Pour Wakefield, cet ajout de la notion de préjudice vient pallier certains problèmes d'une approche purement naturaliste, comme celle de Boorse.

Bien qu'il l'ait peu discutée, Wakefield ajoute deux clarifications à son analyse de la notion de préjudice. D'abord, selon Wakefield, le préjudice doit être causé par la dysfonction (Wakefield, 1997, p. 661). Cette précision implique que le fait de dévaloriser une condition ne peut pas faire de cette condition un trouble si aucune dysfonction n'est présente (Wakefield, 2005, p. 89; voir aussi Wakefield, 2007, Wakefield et Conrad, 2019). Une société peut donc se tromper en identifiant une condition comme un trouble mental. Cette erreur survient lorsque la société considère une condition comme pathologique sans qu'une dysfonction sous-jacente soit postulée. Avec cette précision, il est possible d'exclure des cas comme la drapétomanie du domaine du pathologique (Wakefield, 2007, p. 155). Qui plus est, Wakefield spécifie que le préjudice doit correspondre à des valeurs sociales et culturelles, et non pas individuelles. Par exemple, si une dysfonction liée à la lecture est présente chez un individu, et si l'individu se trouve dans une société qui valorise la lecture, alors il est

possible d'inférer la présence d'un trouble mental, même si l'individu lui-même ne valorise pas la lecture (Wakefield, 2005, p. 89). La réponse attendue à la présence d'un trouble mental (p. ex. : traitement ou non) peut impliquer les valeurs individuelles, mais le diagnostic en lui-même doit relever des valeurs sociales et culturelles. Autrement dit, dire qu'un individu souffre d'un trouble mental nécessite la présence d'une dysfonction et de conséquences entraînées par cette dysfonction qui sont jugées négatives par la société/culture. Or, la décision de suivre un traitement ou non peut relever des valeurs individuelles.

Plus récemment, Wakefield a pointé quelques limites de la DP par rapport au critère de préjudice tel qu'il est décrit dans ses articles des années 1990 et 2000 :

To this extent, my (1992) claim that harm is judged by social values was overly simplistic. For example, imagine that runaway slaves and Soviet dissidents (both claimed by respective social authorities to be disordered) had minor brain dysfunctions that made them less tolerant of oppression and more freedom-aspiring than others. These groups' actions were socially judged as harmful by their societies, potentially justifying a disorder diagnosis if dysfunctions did exist (Wakefield, 2013, p. 1).

Ainsi, Wakefield maintient toujours que le préjudice doit faire référence aux normes de la société, mais ce ne doit pas être uniquement une réponse normative immédiate. Le fait qu'une société juge une dysfonction comme préjudiciable à un moment donné n'entraîne pas nécessairement que ce jugement est réfléchi et légitime. Dans un autre article, Wakefield précise :

Wakefield, although somewhat ambiguously, has located the context of the evaluation of harm in 'cultural standards' [...], not necessarily in actual current attitudes. Wakefield has recently clarified that in the intended sense, social values or standards are not synonymous with the attitudes or opinions that predominate at any given moment. Rather, they are part of a cultural value system that has a complex multilayered structure and that is open to critical

scrutiny and revision in the course of a dialectic about which of a culture's many often-conflicting value commitments are most basic, how to adjudicate between competing values, whether some seeming values are really just rationalisations of unjust power or blind prejudice, and how changing circumstances should alter these judgements (Wakefield et Conrad, 2019, p. 1).

Pour déterminer si une dysfonction entraîne un préjudice, il ne s'agit donc pas de se baser sur les normes communément partagées d'une société ou d'une culture, mais de se baser sur des normes sociales et culturelles « réfléchies ». Or à ce moment-ci, la manière de parvenir à un « bon » jugement sur le préjudice n'est toujours pas clarifiée dans les articles de Wakefield (je reviens sur cet enjeu dans la section 3.3.4).

3.3 Présentation des critiques de la DP

La DP a eu énormément d'influence en philosophie de la psychiatrie. L'intérêt principal de cette approche est d'avoir tenté d'allier les intuitions de l'approche purement naturaliste et de l'approche purement constructiviste par les notions de dysfonction et de préjudice. Spitzer a d'ailleurs affirmé que la DP était une amélioration considérable par rapport aux autres approches qui ont tenté de définir le trouble mental, incluant la sienne (Spitzer, 1999, p. 431). Qui plus est, plusieurs auteurs ont soutenu que le choix de la théorie de la fonction sélectionnée par rapport à l'approche « orientée vers des buts » était un argument en faveur de la théorie de Wakefield par rapport à Boorse (p. ex. : Powell et Scarffe, 2019b, p. 1172-1173). Malgré que le travail de Wakefield apparaisse comme une contribution majeure en philosophie de la psychiatrie, de nombreuses voix critiques se sont élevées contre cette définition du

trouble mental dans les dernières décennies (p. ex. : Aragona, 2009; Bingham et Banner, 2014; Bolton, 2008, p. 126-159; Demazeux, 2015; Faucher et Forest, 2021; Fulford, 1999; Gold et Kirmayer, 2007, p. 165; Houts, 2001; Jablensky, 2007, p. 157-158; Kirmayer et Young, 1999; Lilienfeld et Marino, 1995, 1999; McNally, 2001; Murphy et Woolfolk, 2000; Richters et Hinshaw, 1999; Sadler et Agich, 1995; Schwartz, 2007). Dans cette section, je vais donc passer en revue quatre critiques qui ont visé la DP : le choix de la théorie de la fonction (3.3.1), le problème de la démarcation (3.3.2), les implications pratiques de la DP (3.3.3) et le manque de spécification de la notion de préjudice (3.3.4).

3.3.1 Critiques de l'approche évolutionniste

L'un des problèmes majeurs de la théorie de la DP est le cadre évolutionniste dans lequel s'ancre la notion de dysfonction. Une première limite de ce cadre est le manque de connaissances par rapport à l'histoire évolutive de l'esprit humain. Comme le rapporte McNally :

Because the phylogenetic trajectory of human cognition is difficult, if not impossible, to ascertain, and because establishing that current traits enjoy stabilizing selection is almost as difficult, it is unclear how an evolutionary approach can undergird abnormal psychology in any substantive way (McNally 2020).

Bien qu'il soit difficile de douter que l'esprit humain ait évolué en fonction de pressions de sélection passées, et qu'ainsi certains traits et mécanismes mentaux soient le résultat de la sélection naturelle, l'histoire évolutive exacte de l'esprit humain est difficilement

connaissable. En effet, l'environnement ancestral au sein duquel l'esprit humain a évolué n'est pas parfaitement connu, et les effets bénéfiques de certains mécanismes adaptatifs dans ces environnements encore moins. Les hypothèses sur les fonctions sélectionnées de l'esprit humain sont donc, pour la plupart, spéculatives (voir aussi, entre autres, Aragona, 2009; Bingham et Banner, 2014; Demazeux, 2015; Faucher, 2021; Schramme, 2010 pour des arguments similaires).

Wakefield semble accepter ce point. En effet, l'idée de la boîte noire essentialiste qu'il défend est précisément de soutenir que les fonctions exactes de l'esprit humain n'ont pas besoin d'être connues pour que l'inférence d'une dysfonction soit légitime. La boîte noire entraîne que l'on utilise nos intuitions à la place des faits scientifiques manquants pour juger des cas litigieux (Wakefield, 1997a, p. 256). Pour Wakefield, la limite actuelle des connaissances sur l'évolution de l'esprit humain n'affecte pas le contenu de la définition du concept de trouble mental (en tant que dysfonction préjudiciable). Selon lui, l'enjeu épistémologique d'identifier les dysfonctions serait différent de celui d'analyser la signification du concept de trouble mental (Wakefield, 2001, p. 349). Cette défense apparaît faible. En effet, bien qu'il soit vrai qu'une partie du projet de Wakefield consiste à décrire nos intuitions par rapport à la signification du trouble mental par la méthode de l'analyse conceptuelle, et ainsi parvenir à une définition du trouble mental, cette méthode est employée pour résoudre les controverses en psychiatrie par rapport à différents diagnostics litigieux (Wakefield, 1992a). L'analyse de Wakefield, au-delà de son caractère descriptif par rapport à nos intuitions, doit aussi pouvoir nous aider à distinguer les conditions réellement pathologiques de conditions socialement dévalorisées, mais qui ne sont pas des troubles mentaux. Ici, c'est la notion de dysfonction, comprise en termes évolutionnistes, qui devait permettre de départager de tels cas. Ainsi, ce n'est pas tout de décrire nos intuitions par rapport au concept de trouble mental en termes de dysfonction préjudiciable. Il faut aussi que le résultat de cette analyse puisse être applicable pour trancher certaines controverses. Il est peut-être vrai que la définition de la DP décrit bien les intuitions des experts et des cliniciens

sur ce qu'est un trouble mental. Or, tant et aussi longtemps que le concept de dysfonction n'a pas de contenu empirique (c'est-à-dire que les théories évolutionnistes ne sont pas plus concluantes au regard des fonctions mentales sélectionnées), il est difficile de voir comment juger des catégories diagnostiques particulières telles que le trouble oppositionnel ou le trouble dysphorique prémenstruel, pour lesquelles nous ne savons pas si une dysfonction est présente ou non. Cette limite de la DP apparaît d'autant plus surprenante qu'elle est similaire à la critique qu'avait adressée Wakefield aux approches purement constructivistes. Leur incapacité à distinguer des cas socialement dévalorisés, mais qui ne sont pas des troubles mentaux les rendait insatisfaisantes (voir la section 3.1.2). Bien que Wakefield identifie la notion de dysfonction comme le critère pouvant départager de tels cas, tant que cette notion est limitée par le manque de connaissance empirique par rapport aux réelles fonctions normales de l'esprit humain, la théorie de Wakefield est elle aussi insatisfaisante. Dans l'état actuel des connaissances, la DP demeure incapable de départager de tels cas, à l'image des constructivistes.

Dans une perspective similaire, certains auteurs ont affirmé qu'au-delà du manque de connaissances certaines quant à l'histoire évolutive passée, il existait également plusieurs types d'explication quant aux causes des troubles mentaux dans les écrits en psychologie et en psychiatrie évolutionniste, et qu'aucun consensus n'existait pour le moment quant à la bonne explication ou aux bonnes explications à adopter. Bien que la théorie de Wakefield fasse partie de ces explications, d'autres explications sont possibles (p. ex. : Bolton, 2008, p. 76-82; Cosmide et Tooby, 1999; Murphy, 2005; Richters et Hinshaw, 1999; voir également Murphy et Stich, 2000 pour une présentation générale de ces disciplines et une discussion critique). Muprhy (2005) dresse le portrait de trois grandes théories visant à expliquer le trouble mental en psychologie évolutionniste. Les troubles mentaux pourraient être expliqués de trois façons (au moins) :

- 1) En termes de *bris*, où des systèmes évolués échouent à réaliser leurs fonctions sélectionnées. En ce sens, une dysfonction serait une dérégulation d'un mécanisme fonctionnel sélectionné par sélection naturelle.
- 2) En termes de *décalage*, où les systèmes évolués performant, dans un environnement nouveau, leurs fonctions sélectionnées dans un environnement ancestral. Ce qui était adaptatif dans un environnement passé ne l'est plus, ce qui explique le trouble.
- 3) En termes de *persistance*, où certains troubles mentaux seraient le résultat de structures de personnalité évoluées et impliqueraient la réalisation de fonctions sélectionnées. En ce sens, le trouble mental serait la fonction sélectionnée.

Alors que la première explication correspond à la conception de Wakefield du trouble mental en tant que déviation par rapport à une fonction sélectionnée, Murphy développe davantage les points 2 et 3. Dans cette section, je vais me concentrer sur sa discussion des troubles de l'anxiété tels que les phobies pour illustrer la seconde explication en jeu et ses problèmes³. J'utilise cette discussion puisque pour Murphy, il s'agit de l'une des explications les plus convaincantes pour expliquer un trouble mental dans les écrits en psychiatrie évolutionniste. L'idée sera de montrer que même pour les hypothèses les plus solides, des incertitudes importantes demeurent et des défis épistémologiques de taille se présentent.

³ Murphy discute également le trouble de la dépression majeure en lien avec l'explication 2) (2005, p. 753-757) et la personnalité antisociale (telle que la psychopathie) en lien avec l'explication 3) (2005, p. 757-761). D'autres explications sont également possibles, telles que l'explication de la stratégie d'adaptation (lorsque certains systèmes ne fonctionnent pas comme prévu, des mécanismes alternatifs peuvent être utilisés pour répondre à certains problèmes de l'environnement) et l'explication de l'apprentissage qui mène à un comportement maladaptatif (l'apprentissage est une fonction sélectionnée qui peut parfois permettre de développer des comportements dysfonctionnels non sélectionnés dans de nouveaux environnements), voir Bolton, 2008, p. 76-82, pour une discussion.

Les phobies se caractérisent par une peur ou une anxiété marquée causée par certains objets ou certaines situations, qui causent des préjudices à l'individu en termes de détresse ou d'incapacité dans la vie courante (APA, 2013, p. 197). Selon Murphy (2005), l'une des hypothèses avancées par la psychologie évolutionniste pour expliquer les phobies fait appel au décalage. Dans cette perspective, les personnes atteintes de phobies seraient extrêmement sensibles à la détection de certaines menaces, comme le fait d'être placé dans des environnements spécifiques (p. ex. : en hauteur ou devant certains animaux comme les serpents, les araignées ou les larves). Cette sensibilité, dans un environnement ancestral passé dangereux, aurait conféré un avantage évolutif aux porteurs du trait. En effet, les personnes apeurées par ce type de danger et portées à les fuir auraient survécu davantage et auraient laissé davantage de descendants porteurs de ce trait. Il est donc possible d'imaginer que certains systèmes cognitifs aient évolué pour permettre une réponse à des situations dangereuses spécifiques dans les environnements ancestraux, telles que les systèmes cognitifs de détection de la menace et de production de la peur et de l'anxiété. Or, nos environnements modernes ne présentant plus ce type de dangers de manière aussi aigüe, une sensibilité toujours extrême à ce type de menaces devient problématique (p. ex. : les personnes avec une phobie des hauteurs ne peuvent pas prendre l'avion ou emprunter certains escaliers, ce qui affecte leur qualité de vie). Il y aurait donc un décalage entre l'environnement passé et l'environnement présent, expliquant le trouble de la phobie. La phobie serait donc l'expression d'une fonction normale sélectionnée dans un environnement passé dangereux, mais qui perd son avantage dans l'environnement présent, pouvant même causer des préjudices graves aux individus avec une hypersensibilité aux stimuli dangereux.

Une première implication de ce type d'explication est que si elle est vraie, des conditions généralement considérées comme des troubles mentaux perdront leur statut pathologique. Cela entraînerait un révisionnisme peut-être plus grand que ce que Wakefield avait imaginé. En effet, si les phobies sont en réalité des fonctions normales

sélectionnées, mais réalisées dans un environnement nouveau où la fonction ne donne plus d'avantage évolutif (voire cause un préjudice), alors suivant Wakefield, les phobies ne devraient plus être considérées comme des troubles mentaux. Cette conclusion a quelque chose de contre-intuitif, d'autant plus que les phobies ne semblent pas être des cas isolés de traits fonctionnant dans des environnements nouveaux (p. ex. : psychopathie, sociopathie, dépression, trouble obsessionnel compulsif, etc., voir aussi Cooper, 2005, p. 19-21 pour une discussion). Un autre problème avec l'explication du décalage (et les autres types d'explication évolutionnistes) est que les éléments de preuve en faveur de l'une ou l'autre de ces hypothèses ne sont souvent pas concluants. Par exemple, bien que pour les phobies, l'hypothèse du décalage soit raisonnable, le trouble pourrait aussi être expliqué par d'autres mécanismes. À cet effet, Murphy écrit :

Depending on how one reads some central ideas of the theory, it can be read as a breakdown-type explanation (if anxiety-producing mechanisms are presumed to be dysfunctional); as a mismatch explanation (if we assume phenotypic variation rather than dysfunction but lack of adaptiveness in the modern environment); or as a persistence explanation (if we assume, as I did at the end, continued utility in some environments) (Murphy, 2005, p. 752).

En ce sens, les phobies pourraient aussi être expliquées comme un dysfonctionnement d'un module sélectionné (comme un bris). Dans cette perspective, une phobie pourrait donc être l'expression de capacités dysfonctionnelles à traiter l'information correctement dans un environnement complexe et à ajuster l'interprétation des stimuli en fonction d'informations supplémentaires, puis à déclencher la bonne réponse face à cet ensemble d'informations. Or selon Murphy, il est très difficile de trancher entre ces deux types d'explication (voir aussi Faucher et Blanchette, 2011 pour une discussion portant sur les éléments de preuve non concluants, voire incompatibles, concernant certaines hypothèses sur les modules évolués de la peur et l'explication évolutionniste des phobies). Ces controverses et l'absence d'éléments de preuve concluants sont caractéristiques de la psychologie et de la psychiatrie évolutionnistes. Ils plaident

encore une fois en faveur de l'idée selon laquelle il apparaît extrêmement difficile d'appliquer la théorie de l'évolution à la psychopathologie dans l'état actuel de nos connaissances.

Rappelons que selon Wakefield, il n'est pas nécessaire d'avoir des éléments de preuve solides quant à la présence d'une dysfonction sous-jacente pour pouvoir affirmer qu'il s'agisse d'un trouble mental. Des hypothèses et inférences réfléchies suffisent. Ainsi, bien que la notion de dysfonction soit en principe scientifique et indépendante des valeurs parce qu'elle est expliquée par la théorie de l'évolution par sélection naturelle, en l'absence de connaissances abouties sur l'histoire évolutive passée de l'être humain, nos intuitions deviennent les seuls juges du statut pathologique d'une condition. Or ces intuitions peuvent, elles, se rapporter davantage à la notion de préjudice ou à des conceptions de la normalité préconçues. Ces intuitions peuvent donc être influencées par les valeurs, contrevenant à l'ambition de Wakefield de fournir une notion de dysfonction indépendante des valeurs qui servirait à départager des cas controversés. Un argument similaire a été offert par De Block et Sholl (2021). Selon les auteurs, dans la pratique, il est très difficile d'identifier ce qu'est une dysfonction sans faire appel aux valeurs. Selon De Block et Sholl, bien que pour la DP, en théorie, la sélection naturelle soit le critère de démarcation entre le normal et le pathologique (conjointement avec le critère du préjudice), dans la pratique, ce sont plutôt les jugements de valeur par rapport à ce qui est préjudiciable qui nous permet de départager les conditions pathologiques des conditions saines. En l'état actuel des connaissances sur les structures évoluées de l'esprit humain, la DP ne repose pas sur une théorie scientifique solide et bien établie.

3.3.2 Le problème de la démarcation

Schwartz (2007) a défendu que le problème de la démarcation, qui était un problème de taille pour la BST de Boorse (voir le chapitre 2), s'appliquait aussi à la DP (voir aussi De Block et Sholl, 2021; Gold et Kirmayer, 2007 pour un argument similaire). Selon cette perspective, dans la théorie de la fonction comme effet sélectionné adoptée par Wakefield, il n'est pas possible de tracer la ligne entre les niveaux normaux et pathologiques de performance de fonctions sans faire appel aux valeurs. La section précédente a mis en évidence la difficulté d'identifier les fonctions normales sélectionnées dans un cadre évolutionniste. Schwartz montre que même si ce cadre permettait d'identifier les fonctions normales sélectionnées, il faudrait également établir quel niveau de performance de la fonction est normal, et à partir de quand un niveau devient pathologique. L'enjeu est donc de pouvoir identifier un seuil de performance pathologique malgré la grande diversité de niveaux de performance pour une fonction donnée. Il faut en effet pouvoir 1) établir que la fonction normale du cœur est de pomper le sang et 2) établir le niveau de performance normal selon lequel le cœur pompe le sang, selon une fraction d'éjection (FE) particulière). Schwartz écrit :

Even among healthy 70-year-old men, for instance, there is a large range in the ability of hearts to pump. People who run marathons have very efficient hearts, reflected in their EF and other features, while those who rarely exercise have much less efficient hearts. But there are boundaries to this normal range : below a certain level of functioning, the heart counts as abnormal. Making sense of “normal variation” requires solving the line-drawing problem as well (Schwartz, 2007, p. 366).

L'argument de Schwartz est alors que bien que la théorie de la fonction de Wakefield permette d'établir les fonctions normales sans faire appel à la notion de préjudice, elle

ne permet pas de démarquer les niveaux de performance sains et les niveaux de performance pathologiques. Autrement dit, bien qu'il soit possible pour Wakefield d'affirmer qu'un cœur qui ne pompe pas le sang est dysfonctionnel, un cœur qui pomperait le sang selon un niveau de performance très faible ne pourrait pas être dit dysfonctionnel, parce que performant tout de même sa fonction (voir aussi Bolton, 2008, p. 146 pour un argument similaire). Donc, dans la pratique, le critère de démarcation pour établir le niveau de performance normal est le préjudice.

Notons d'abord, suivant Schwartz (2007, p. 370), que Wakefield défend que la sélection naturelle ne sélectionne pas uniquement des fonctions, mais aussi des niveaux de performance de fonctions. Par exemple, Wakefield affirme que dans les cas d'attaque de panique, la sélection aurait bel et bien sélectionné des mécanismes de peur et d'anxiété, mais que le niveau de peur et d'anxiété impliqué dans les attaques de panique serait trop élevé et n'aurait pas été sélectionné (voir Wakefield, 1999b, p. 379). Or pour Schwartz, ce type d'affirmation est difficilement démontrable à partir du cadre théorique adopté par Wakefield. Supposons une situation où différents niveaux de performance d'un mécanisme sont associés à différents niveaux de *fitness* (compris comme la capacité à survivre et à se reproduire des individus). Lequel ou lesquels de ces niveaux ont été sélectionnés? Il apparaît encore plus difficile de déterminer le niveau de performance sélectionné que la fonction normale sélectionnée. Ainsi, en l'absence de connaissances sur le niveau de performance sélectionné, ce seront les conséquences positives ou négatives du niveau de performance d'une fonction sur la vie de l'individu qui permettront de déterminer le seuil de performance pathologique. Si à partir d'un certain seuil, les conséquences négatives sur les individus sont importantes, ce niveau de performance sera dit pathologique. Encore une fois, dans la pratique, les valeurs entrent en jeu pour démarquer le normal du pathologique étant donné que le cadre évolutionniste ne fournit pas ce type d'informations.

Avant de clore cette section, une précision s'impose. Wakefield a répondu à un argument similaire avancé par Lilienfeld et Marino (1999), qui défendaient que puisque la distinction entre ce qui est un effet sélectionné et ce qui ne l'est pas est vague, le concept de dysfonction est arbitraire. La réponse de Wakefield est que cette distinction, parce que fondée dans la théorie de l'évolution, est scientifique, et donc non arbitraire. Mais la critique de Schwartz est plus subtile et plus incisive : puisque la ligne entre le normal et le pathologique en termes de niveau de performance est tracée en raison d'une prise de décision, et non pas en « lisant la nature », des valeurs sont présentes dans le concept de dysfonction lorsque vient le temps de déterminer le niveau normal de performance d'une fonction. Cela ne veut pas dire que ces décisions sont arbitraires, mais bien que ces décisions sont motivées par autre chose que des faits empiriques. La notion de dysfonction est donc à tout le moins partiellement influencée par des valeurs, et non pas purement indépendante des valeurs. Cela vient remettre en question la distinction stricte que dresse Wakefield entre, d'un côté, le préjudice, influencé par les valeurs, et de l'autre, la dysfonction, purement factuelle et scientifique. L'argument de Schwartz semble convaincant et plaide en faveur d'une remise en question de cette distinction.

3.3.3 Critique des implications pratiques de la DP

Une critique intéressante, mais davantage liée aux implications pratiques de l'application de la DP aux cas actuellement reconnus de troubles mentaux, mérite d'être rapportée. Cette critique est mise de l'avant par Bolton (2008). L'idée de Bolton est la suivante : si au moins certains types d'explication évolutionniste des troubles mentaux présentés dans la section précédente sont vrais (ce qui reste à démontrer

empiriquement), il est possible de penser que certaines des conditions actuellement considérées comme des troubles mentaux ne soient pas explicables en termes de dysfonction. Cette possibilité contreviendrait à la définition donnée par Wakefield, puisque bien que ces conditions puissent causer des préjudices aux individus, aucune dysfonction (en termes évolutionnistes) ne serait présente. Wakefield (2000, p. 259, 2011, p. 152) semble concéder ce point, et affirme que dans cette éventualité, nous serions contraints de changer nos jugements sur ces conditions : ce qui était considéré comme un trouble mental ne devrait plus l'être, puisqu'aucune dysfonction n'est présente. Rappelons que dans l'état actuel, l'un des problèmes est la difficulté d'identifier une dysfonction (Bolton, 2008, p. 131-134). Or même si nous avons toutes les connaissances nécessaires pour identifier les dysfonctions, un autre problème surgirait : que faire des catégories diagnostiques qui ne sont pas des dysfonctions, mais qui causent tout de même de la souffrance, de la détresse, un handicap ou qui entraînent un risque élevé de conséquences négatives pour l'individu (Bolton, 2008, p. 135-139)? Selon Bolton, il est possible de penser que ces conditions devraient demeurer dans le DSM, même si elles ne sont pas associées à des dysfonctions. Elles pourraient être qualifiées par une autre étiquette que « trouble mental », mais demeurer dans le DSM parce qu'elles demanderaient toujours une attention et une prise en charge cliniques. Dans la section précédente, j'ai discuté le cas des phobies dans le cadre explicatif du décalage. Si cette explication est vraie, alors les phobies ne sont pas des cas de dysfonctions, mais des cas de fonctions sélectionnées dans des environnements ancestraux qui deviennent préjudiciables dans des environnements nouveaux. Suivant Wakefield, les phobies ne devraient plus être considérées comme des troubles mentaux. Or il est reconnu que les phobies, dans leur manifestation extrême, peuvent causer d'importants préjudices aux individus. Qu'advierait-il des individus souffrant de phobies si ces conditions n'étaient pas considérées comme des troubles mentaux? Le problème ne s'arrête pas aux phobies : il est possible d'imaginer que plusieurs conditions qui ne sont pas associées à des dysfonctions puissent être accompagnées de préjudices importants (p. ex. : supposons que le trouble dysphorique prémenstruel ne

soit pas causé par une dysfonction, il n'en demeure pas moins que de nombreuses femmes souffrent de cette condition, qui affecte leur qualité de vie de manière importante). Cet exemple démontre que le fait qu'une condition préjudiciable ne soit pas causée par une dysfonction ne semble pas affecter l'aspect clinique du diagnostic, c'est-à-dire l'importance d'une prise en charge thérapeutique pour plusieurs types de conditions préjudiciables, qu'une dysfonction soit présente ou non.

Bolton (2008, p. 139) fait alors l'hypothèse que si nous avons les connaissances évolutionnistes nécessaires pour identifier les dysfonctions, et si plusieurs catégories diagnostiques du DSM n'étaient pas explicables par des dysfonctions, alors le DSM serait découpé en deux grandes catégories : les troubles mentaux (pour lesquels des dysfonctions sont associées) et les « conditions psychiatriques », pour lesquels aucune dysfonction ne serait connue, mais qui seraient toujours associées à des préjudices. Comme le soutient Bolton, il n'est pas clair que la distinction entre ces deux grandes classes de conditions incluses dans le DSM apporte réellement quelque chose à la psychiatrie, puisque d'un point de vue clinique, la réponse devrait être la même : une prise en charge thérapeutique, peu importe si une dysfonction est présente ou non. Qui plus est, il faudrait se poser la question : quel type de condition psychiatrique qui n'est pas sous-tendue par une dysfonction doit être inclus dans le DSM en tant que « condition psychiatrique »? La DP ne donne pas de réponse à cette question.

3.3.4 Critique de la notion de préjudice

Alors que les sections précédentes visaient davantage la notion de dysfonction et ses prétentions à demeurer indépendante des valeurs sociales et culturelles, d'autres

critiques ont visé la notion de préjudice, soit la seconde composante essentielle identifiée par la DP. C'est ce que font De Block et Sholl (2021) : selon les deux auteurs, le choix de Wakefield de prioriser les valeurs sociales et culturelles, et non pas les valeurs individuelles, n'est pas justifié. Certaines conditions pourraient être perçues par la société comme préjudiciables, mais pas par les individus atteints de ces conditions. Par exemple, De Block et Sholl soutiennent qu'il paraît difficile de défendre qu'un individu atteint de dyslexie, mais qui ne fait pas l'expérience individuelle d'un préjudice (en termes de souffrance ou de détresse, par exemple), soit classé comme ayant un trouble mental, bien que les sociétés lettrées dévalorisent les difficultés d'apprentissage. Or c'est pourtant l'implication de l'analyse de Wakefield, qui situe les valeurs pertinentes pour le concept de trouble mental au niveau de la société, et non pas de l'individu (voir la section 3.2.4). Par ailleurs, De Block et Sholl précisent que les normes sociales ne sont pas interprétées par tout le monde de la même façon, ni acceptées par tous les individus d'une société. Il y a donc fort à parier que plusieurs visions sur une condition particulière cohabiteront dans une société donnée, rendant difficile l'évaluation du degré de préjudice associé par une société à une condition particulière. Ajoutons que la manière selon laquelle il faut identifier si un préjudice est présent (au niveau social) n'est pas explicitée par Wakefield. Devrait-on se fier à nos intuitions ou procéder à des études expérimentales sur les valeurs attribuées à un trouble? Rappelons par ailleurs que Wakefield affirme qu'une société peut se tromper par rapport aux valeurs qu'elle attribue à une condition (Wakefield et Conrad, 2009), mais qu'il n'élabore pas vraiment sur la manière de déterminer s'il s'agit d'une erreur ou plus largement sur quelles sont les « bonnes » et les « mauvaises » valeurs.

Qui plus est, même si l'on accepte cette critique du choix des valeurs sociales et culturelles, De Block et Sholl arguent que situer les valeurs pertinentes au niveau de l'individu implique également son lot de problème. Dans un premier temps, il semble que les valeurs individuelles aient le plus souvent, au moins en partie, une origine sociale. En effet, le fait de ressentir un préjudice sous la forme de stigma social trouve

son origine dans l'environnement social où se situe la personne. Les niveaux individuels et sociaux sont donc interreliés, et possiblement difficilement distinguables. Supposons une société homophobe, un individu pourra ressentir le stigma social contre son orientation sexuelle et en souffrir. L'individu subirait donc individuellement un préjudice, qui trouve son origine au niveau social. Situer les valeurs pertinentes au niveau individuel ne règle donc pas tout. Dans un second temps, De Block et Sholl affirment que certaines conditions s'accompagnent d'anosognosie, soit le fait de ne pas être conscient de sa condition. Il devient alors difficile pour un individu de juger des effets (préjudiciables ou non) de sa condition. Ces deux critiques ne démontrent pas qu'une définition du trouble mental ne doit pas intégrer une composante axiologique, ni à quel niveau devrait être cherchée cette composante. Elles pointent plutôt la difficulté à rendre compte des valeurs qui entrent en jeu dans la définition du trouble mental et les défis qui ne sont pas adressés par Wakefield dans l'état actuel de sa théorie.

Dans la même lignée, Powell et Scarffe (2019a, 2019b) ont récemment défendu que les jugements de valeur pouvaient être rationnellement réfléchis et que cette possibilité n'était pas suffisamment prise en compte dans l'œuvre de Wakefield⁴. Pour Powell et

⁴ Cependant, Powell et Scarffe (2019a, 2019b) adoptent la vision de la dysfonction de Wakefield et la jugent satisfaisante. Par exemple, lorsqu'ils discutent un cas d'étude, celui du diagnostic de stress post-traumatique, ils écrivent : « Although the precise neural mechanisms underlying PTSD are poorly understood, researchers have identified several neurobiological deficits that are strongly indicative of emotional dysregulation. Traumatic stressors result in a number of cognitive affective dysregulations, some of which are common to anxiety disorders while others look to be specific to PTSD. These include exaggerated amygdala and insula activity which modulate fear and startle responses; as well as hypoactivation in brain regions associated with the processing of negative emotions (including the ventromedial prefrontal cortex, rostral and dorsal anterior cingulate cortex, thalamus and the dorsal amygdala), which results in the autonomic blunting associated with episodes of numbing and dissociation that are characteristic of PTSD. Thus, PTSD appears to satisfy the moral and biological criteria of our proposed thickly normative hybrid account of disease. » (Powell et Scarffe, 2019b, p. 583) Or, ce n'est pas parce qu'un phénomène neurophysiologique sous-tend une condition psychiatrique particulière que cette condition est une dysfonction à proprement parler. La couleur des yeux s'explique par des mécanismes particuliers, mais la variation de la couleur des yeux n'implique pas de dysfonction.

Scarffe, le principal problème de l'approche constructiviste et de l'approche hybride est le manque de normativité de leur modèle :

Homosexuality should not be considered a disease because devaluing homosexuality is not rationally justified and causes objective harm. The real problem with the hybrid approach, then, is that whether it classifies a biomedical state as a disease is contingent on mere facts about evolution and social evaluation. Thus, although weakly normative hybrid accounts can make sense of misapplications that do not implicate biological dysfunctions (e.g., masturbation), they do not have the resources to recognize as misapplications biological dysfunctions that are wrongfully or perniciously devalued (Powell et Scarffe, 2019b, p. 1174-1175).

Ici, Powell et Scarffe ciblent l'incapacité de la DP à se prononcer contre l'application de l'étiquette de trouble mental à une condition qui serait une dysfonction et qui serait socialement dévalorisée pour de mauvaises raisons. Cette critique est similaire à celle adressée par Wakefield lui-même aux normativistes (voir la section 3.1.2) : la DP ne serait pas assez normative dans les cas où une société, pour de mauvaises raisons, jugerait des dysfonctions comme préjudiciables (Powell et Scarffe, 2019a, p. 582). Pour les auteurs, en raison de cette lacune, la théorie de la DP doit être modifiée pour pouvoir identifier les cas de dysfonctions qui sont correctement dévalorisés (ou pour lesquels un préjudice est bel et bien présent). Ils proposent alors une version modifiée de la DP, qu'ils appellent la théorie hybride normativement épaisse (*thickly normative hybrid theory*). Dans cette théorie, c'est la justification rationnelle et morale qui peut permettre de juger correctement de la présence ou non d'un préjudice, même si une société particulière considère la présence d'un préjudice à un moment donné. Par la

Par ailleurs, les mécanismes cités par Powell et Scarffe ne sont pas nécessairement conçus comme des dysfonctions dans le cadre théorique sélectionniste : nous n'avons pas les connaissances pour affirmer ce qui a été sélectionné ou non en termes de mécanismes régulateurs des émotions. Ainsi, bien que leur suggestion pour améliorer la DP de Wakefield par rapport à la notion de préjudice soit pertinente, leur conception de la fonction doit être revue.

justification rationnelle morale, le concept de trouble mental peut acquérir de l'objectivité aussi en ce qui concerne le préjudice. À cet effet, ils écrivent :

What uncontroversial things can we say about the 'rational justification' element, given that the methods of moral justification are contested? At the very least, we can say that reasons independent of popularity and tradition must be adduced to justify the evaluation. We can say that these reasons must not be arbitrary or bigoted, that they should not repose on false empirical claims or inscrutable religious beliefs and that they must be subject to critical scrutiny and revision (Powell et Scarffe, 2019a, p. 582).

Les auteurs ajoutent également que le consensus moral dans une société n'est pas nécessairement un signe que les valeurs morales adoptées le sont pour les bonnes raisons (2019a, p. 583). C'est plutôt que les processus de justification morale doivent être les bons (Powell et Scarffe, 2019a, p. 583). Ainsi, selon Powell et Scarffe, les valeurs légitimement contenues dans la définition du trouble mental devraient être identifiées grâce à une justification morale rationnelle pour qu'une conception hybride soit satisfaisante. Dans les chapitres 5 et 6, je tire les implications de cette proposition et je les applique à une conception constructiviste et objective de trouble mental.

Conclusion

Dans ce chapitre, je me suis penchée sur deux approches alternatives à la BST de Boorse : le constructivisme et l'approche hybride. Mon analyse a révélé que bien que le constructivisme soit convaincant dans sa défense de la thèse de l'influence des valeurs sur le concept de trouble mental, il demeure insatisfaisant dans sa forme

actuelle. Plus spécifiquement, c'est son incapacité à justifier l'influence des valeurs sur le concept et à départager les valeurs légitimes et illégitimes qui constitue sa principale limite. L'approche de Wakefield se voulait une solution à ce problème : en posant le critère de la dysfonction (neutre et scientifique), conjointement au critère de préjudice (influencé par les valeurs), l'ambition de Wakefield était de parvenir à fonder la distinction entre le normal et le pathologique sur une base objective et scientifique. L'examen critique de cette proposition me conduit à soutenir que l'ambition de Wakefield échoue, en ce que la DP rencontre des problèmes relativement à ces deux critères. Ancrer la notion de dysfonction dans le cadre de l'évolution naturelle, bien que séduisant à première vue, implique une difficulté pratique à départager les cas controversés en raison de l'état des connaissances actuelles et des implications possiblement problématiques s'il s'avérait que plusieurs troubles mentaux soient en fait le résultat de la sélection naturelle. En ce qui concerne le préjudice, cette notion est sous-développée chez Wakefield. En raison des problèmes caractéristiques des approches naturaliste, constructiviste et hybride, il me paraît important d'effectuer un pas de recul par rapport au débat traditionnel sur la définition du trouble mental et de m'interroger sur certaines intuitions de ces approches. Dans le chapitre suivant, je vais questionner l'un des présupposés de l'approche objectiviste des fonctions naturelles (dont Wakefield et Boorse font partie) : l'idée que seule une notion de dysfonction indépendante des valeurs puisse permettre de dépasser les problèmes du constructivisme.

Références du chapitre 3

- Agich, G. J. (1983). Disease and value : A rejection of the value-neutrality thesis. *Theoretical Medicine*, 4(1), 27-41.
- Agich, G. J. (1997). Toward a pragmatic theory of disease. Dans J. M. Humber et R. F. Almeder (dir.), *What is disease?* (p. 219-246). Totowa, NJ : Humana Press.
- Agich, G. J. (2002). Implications of a pragmatic theory of disease for the DSMs. Dans J. Z. Sadler (dir.), *Descriptions and prescriptions : Values, mental disorders, and the DSMs* (p. 96-113). Baltimore : Johns Hopkins University Press.
- Aragona, M. (2009). The concept of mental disorder and the DSM-V. *Dialogues in Philosophy, Mental and Neuro Sciences*, 2(1), 1-14.
- Aucouturier, V. et Demazeux, S. (2013). The concept of “mental disorder”. Dans H. Carel et R. Cooper (dir.), *Health, illness and disease* (p. 89-106). Londres : Routledge.
- Bingham, R. et Banner, N. (2014). The definition of mental disorder : Evolving but dysfunctional? *Journal of Medical Ethics*, 40(8), 537-542.
- Bolton, D. (2008). *What is mental disorder? An essay in philosophy, science, and values*. New York : Oxford University Press.
- Clouser K. D., Culver C. M. et Gert B. (1981). Malady : A new treatment of disease. *The Hastings Center Report*, 11(3), 29-37.
- Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(2), 263-282.
- Cooper, R. V. (2005). *Classifying madness*. Dordrecht : Springer.

- Cosmides, L. et Tooby, J. (1999). Toward an evolutionary taxonomy of treatable conditions. *Journal of Abnormal Psychology*, 108(3), 453-464.
- De Block, A. et Sholl, J. (2021). Harmless dysfunctions and the problem of normal variation. Dans L. Faucher et D. Forest (dir.), *Defining mental disorders : Jerome Wakefield and his critics* (p. 495-510). Cambridge, MA : MIT Press.
- Demazeux, S. (2015). The function debate and the concept of mental disorder. Dans • P. Huneman, G. Lambert et M. Silberstein (dir.), *Classification, disease and evidence* (p. 63-91). Springer : Dordrecht.
- Engelhardt, H. T. (1974). The disease of masturbation : Values and the concept of disease. *Bulletin of the History of Medicine*, 48(2), 234-248.
- Engelhardt, H. T., Jr. (1975). The concepts of health and disease. Dans H. T. Engelhardt, Jr. et S. F. Spicker (dir.), *Evaluation and explanation in the biological sciences* (p. 125-141). Dordrecht : D. Reidel.
- Ereshefsky, M. (2009). Defining “health” and “disease”. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 40(3), 221-227.
- Faucher, L. et Blanchette, I. (2011). Fearing new dangers : Phobias and the cognitive complexity of human emotions. Dans A. De Block et P. Adriaens (dir.), *Maladapting minds : Philosophy, psychiatry, and evolutionary theory* (p. 33-64). New York : Oxford University Press.
- Fulford, K. W. M. (1999). Nine variations and a coda on the theme of an evolutionary definition of dysfunction. *Journal of Abnormal Psychology*, 108(3), 412-420.
- Glackin, S. (2019). Grounded disease : Constructing the social from the biological in medicine. *The Philosophical Quarterly*, 69, 258-276.
- Gold, I. et Kirmayer, L. J. (2007). Cultural psychiatry on Wakefield’s procrustean bed. *World Psychiatry*, 6(3), 165-166.

- Houts, A. C. (2001). Harmful dysfunction and the search for value neutrality in the definition of mental disorder : Response to Wakefield, part 2. *Behaviour Research and Therapy*, 39(9), 1099-1132.
- Jablensky, A. (2007). Does psychiatry need an overarching concept of “mental disorder”? *World Psychiatry*, 6(3), 157-158.
- Kingma, E. (2016). Situation-specific disease and dispositional function. *The British Journal for the Philosophy of Science*, 67(2), 391-404.
- Kirmayer, L. J. et Young, A. (1999). Culture and context in the evolutionary concept of mental disorder. *Journal of Abnormal Psychology*, 108(3), 446-452.
- Lemoine, M. (2017). *Introduction à la philosophie des sciences médicales*. Paris : Hermann.
- Lilienfeld, S. O. et Marino, L. (1995). Mental disorder as a Roschian concept : A critique of Wakefield’s “harmful dysfunction” analysis. *Journal of Abnormal Psychology*, 104(3), 411-420.
- Lilienfeld, S. O. et Marino, L. (1999). Essentialism revisited : Evolutionary theory and the concept of mental disorder. *Journal of Abnormal Psychology*, 108(3), 400-411.
- McNally, R. J. (2001). On Wakefield’s harmful dysfunction analysis of mental disorder. *Behaviour Research and Therapy*, 39(3), 309-314.
- Megone, C. (1998). Aristotle’s function argument and the concept of mental illness. *Philosophy, Psychiatry, & Psychology*, 5(3), 187-201.
- Megone, C. (2000). Mental illness, human function, and values. *Philosophy, Psychiatry, & Psychology*, 7(1), 45-65.
- Millikan, R. G. (1989). An ambiguity in the notion “function”. *Biology and Philosophy*, 4(2), 172-176.

- Murphy, D. (2005). Can evolution explain insanity? *Biology and Philosophy*, 20(4), 745-766.
- Murphy, D. (2006). *Psychiatry in the scientific image*. Cambridge, MA : MIT Press.
- Murphy, D. (2020). Concepts of disease and health. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/spr2020/entries/health-disease/>
- Murphy, D. et Woolfolk, R. L. (2000). The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, & Psychology*, 7(4), 241-252.
- Murphy, D. et Stich, S. (2000). Darwin in the madhouse : Evolutionary psychology and the classification of mental disorders. Dans P. Carruthers et A. Chamberlain (dir.), *Evolution and the human mind : Modularity, language and meta-cognition* (p. 62-92). Cambridge : Cambridge University Press.
- Neander, K. (1991). Functions as selected effects : The conceptual analyst's defense. *Philosophy of Science*, 58(2), 168-184.
- Nordenfelt, L. (1995). *On the nature of health : An action-theoretic approach*. Dordrecht : Springer.
- Nordenfelt, L. (2001). *Health, science, and ordinary language*. Amsterdam : Rodopi.
- Powell, R. et Scarffe, E. (2019a). Rethinking "disease" : A fresh diagnosis and a new philosophical treatment. *Journal of Medical Ethics*, 45(9), 579-588.
- Powell, R. et Scarffe, E. (2019b). Rehabilitating "disease" : Function, value, and objectivity in medicine. *Philosophy of Science*, 86(5), 1168-1178.
- Richters, J. E. et Hinshaw, S. P. (1999). The abduction of disorder in psychiatry. *Journal of Abnormal Psychology*, 108(3), 438-445.
- Sadler, J. Z. et Agich, G. J. (1995). Diseases, functions, values, and psychiatric classification. *Philosophy, Psychiatry, & Psychology*, 2(3), 219-231.

- Schramme, T. (2010). Can we define mental disorder by using the criterion of mental dysfunction? *Theoretical Medicine and Bioethics*, 31(1), 35-47.
- Schwartz, P. H. (2007). Decision and discovery in defining “disease”. Dans H. Kincaid et J. McKittrick (dir.), *Establishing medical reality : Essays in the metaphysics and epistemology of biomedical science* (p. 47-63). Heidelberg : Springer Netherlands.
- Spitzer, R. L. (1999). Harmful dysfunction and the DSM definition of mental disorder. *Journal of Abnormal Psychology*, 108(3), 430-432.
- Varga, S. (2015). *Naturalism, interpretation, and mental disorder*. New York : Oxford University Press.
- Wakefield, J. C. (1992a). Disorder as harmful dysfunction : A conceptual critique of DSM-III-R’s definition of mental disorder. *Psychological Review*, 99(2), 232-247.
- Wakefield, J. C. (1992b). The concept of mental disorder : On the boundary between biological facts and social values. *American Psychologist*, 47(3), 373-388.
- Wakefield, J. C. (1997). Diagnosing DSM-IV–Part II : Eysenck (1986) and the essentialist fallacy. *Behaviour Research and Therapy*, 35(7), 651-665.
- Wakefield, J. C. (1999a). Mental disorder as a black box essentialist concept. *Journal of Abnormal Psychology*, 108(3), 465-472.
- Wakefield, J. C. (1999b). Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology*, 108(3), 374-399.
- Wakefield, J. C. (2001). Evolutionary history versus current causal role in the definition of disorder : Reply to McNally. *Behaviour Research and Therapy*, 39(3), 347-366.
- Wakefield, J. C. (2003). Dysfunction as a factual component of disorder : Reply to Houts, Part 2. *Behaviour Research and Therapy*, 41(8), 969-990.

- Wakefield, J. C. (2004). The myth of open concepts : Meehl's analysis of construct meaning versus black box essentialism. *Applied and Preventive Psychology*, 11(1), 77-82.
- Wakefield, J. C. (2005). On winking at the facts, and losing one's Hare : Value pluralism and the harmful dysfunction analysis. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, 4(2), 88-89.
- Wakefield, J. C. (2007). The concept of mental disorder : Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, 6(3), 149-156.
- Wakefield, J. C. (2013). Addiction, the concept of disorder, and pathways to harm : Comment on Levy. *Frontiers in Psychiatry*, 4(34).
doi.org/10.3389/fpsy.2013.00034
- Wakefield, J. C. (2015). Biological function and dysfunction : Conceptual foundations of evolutionary psychopathology. Dans David M. Buss (dir.), *The Handbook of Evolutionary Psychology*.
doi:10.1002/9781119125563.evpsych242
- Wakefield, J. C. et Conrad, J. A. (2019). Does the harm component of the harmful dysfunction analysis need rethinking? Reply to Powell and Scarffe. *Journal of Medical Ethics*, 45(9), 594-596.
- Whitbeck, C. (1981). A theory of health. Dans A. L. Caplan, H. T. Engelhardt et J. J. McCartney (dir.), *Concepts of health and disease : Interdisciplinary perspectives* (p. 611-626). Reading, VA : Addison-Wesley.

CHAPITRE 4

L'IDÉAL DE LA SCIENCE INDÉPENDANTE DES VALEURS ET LA DÉFINITION DU TROUBLE MENTAL

Introduction

Comme je l'ai présenté dans les chapitres précédents, la thèse de la neutralité du concept de trouble mental grâce à la notion de dysfonction est insoutenable. Or au regard des problèmes rencontrés par les potentielles approches alternatives, soit les différentes formes de constructivisme, il peut être intéressant de se pencher sur les motivations qui ont poussé les représentants de l'approche objectiviste des fonctions naturelles à formuler une définition neutre du trouble mental. Je prétends que comprendre les motivations sous-jacentes à cette approche peut nous aider à dépasser le cul-de-sac dans lequel se trouve ce débat traditionnel en philosophie de la psychiatrie. Si ces motivations sous-jacentes sont inadéquates, alors un espace peut être ouvert pour penser un nouveau cadre théorique dans lequel la place et le rôle des valeurs dans la définition du trouble mental pourront être circonscrits.

Dans la section 4.1, je vais lier l'approche objectiviste des fonctions naturelles à un idéal classique en philosophie des sciences, l'indépendance des valeurs. Dans la section 4.2, j'exposerai plus en détail cet idéal et la relation qu'il a entretenue avec la notion

d'objectivité. Les deux sections suivantes seront consacrées aux critiques qui ont émergé vis-à-vis l'idéal de l'indépendance des valeurs, d'abord en philosophie féministe des sciences (section 4.3), puis en philosophie plus générale des sciences (section 4.4). Dans la section 4.5, j'exposerai le cadre théorique dans lequel s'inscrit l'une des critiques les plus influentes de l'indépendance des valeurs, l'empirisme contextualiste critique, et la critique méthodologique telle que développée par Helen Longino (1990, 2002). Dans la section 4.6, j'exposerai le nouvel idéal qui découle de la critique méthodologique et du cadre de l'empirisme contextuel critique, l'objectivité sociale (OS). Pour finir, dans la section 4.7, je vais démontrer comment l'OS peut être pertinente pour penser la définition du trouble mental en psychiatrie.

4.1 L'approche objectiviste des fonctions naturelles et l'idéal de la science indépendante des valeurs

Selon les défenseurs de la NFO, seulement une définition objective (totalement pour Boorse, ou partiellement pour Wakefield) du trouble mental pouvait répondre aux critiques des auteurs du mouvement antipsychiatrique. Autrement dit, une définition indépendante des valeurs sociales et culturelles pourrait aider à répondre à ces attaques. C'est par l'inclusion d'une notion de dysfonction neutre que la NFO a pensé pouvoir résoudre les difficultés auxquelles faisait face la psychiatrie. Selon Boorse et Wakefield, si la notion de dysfonction est neutre sur le plan des valeurs, alors il devient possible de soutenir que le concept de trouble mental n'est pas utilisé pour contrôler les comportements socialement déviants, mais plutôt pour désigner un phénomène réel de déviation psychophysiologique. C'est en ce sens que Boorse écrit : « *the value-free scientific disease concept remains as a bedrock requirement to block the subversion of*

medicine by political rhetoric or normative eccentricity » (Boorse, 1997, p. 100, les italiques sont les miennes). Dans le même esprit, Wakefield affirme « [o]ne needs *an additional factual component* to distinguish disorders from the many other negative mental conditions not considered disorders, such as ignorance, lack of skill, lack of talent, low intelligence, illiteracy, criminality, bad manners, foolishness, and moral weakness » (Wakefield 2007, p. 151, les italiques sont les miennes; voir aussi Wakefield, 1992a, p. 381). Une notion de dysfonction qui représenterait la nature telle qu'elle est, indépendamment des valeurs sociales et culturelles, pourrait permettre de départager de manière non arbitraire les cas controversés en psychiatrie. Le concept de santé mentale serait alors un vrai concept scientifique et la psychiatrie pourrait baser sa pratique clinique sur cette fondation empirique. La notion de dysfonction, parce qu'indépendante des valeurs, serait donc ce qui donnerait à la psychiatrie son caractère scientifique en ce qui concerne la définition du trouble mental.

Cette idée que la notion de dysfonction permettrait de démarquer objectivement le normal du pathologique se retrouve également chez d'autres auteurs en philosophie de la psychiatrie et en philosophie de la médecine tels que Tyrer et Steinberg (2003, p. 8) et Kendell (1986). Tyrer et Steinberg affirment en effet que l'approche objectiviste des fonctions naturelles permet de contraindre la médicalisation des problèmes normaux de vie et les traitements médicaux qui s'en suivent en fournissant des critères objectifs pour départager le normal du pathologique. L'opposition stricte entre ce qui serait de l'ordre du sociopolitique et ce qui serait de l'ordre du scientifique se lit aussi chez Kendell, un autre défenseur de l'approche naturaliste : « The most fundamental issue, and also the most contentious one, is whether disease and illness are normative concepts based on value judgments, or whether they are value-free scientific terms; in other words, whether they are biomedical terms or sociopolitical ones » (Kendell, 1986, p. 23). L'idée qu'une notion scientifique indépendante des valeurs soit la solution aux controverses se manifestant sur le plan social semble donc communément partagée au sein de l'approche naturaliste.

Pour plus de clarté, il peut être opportun de reconstruire l'argument de la NFO comme suit :

- 1) Si la notion de trouble mental est un concept scientifique, alors la psychiatrie peut répondre aux critiques qui ont attaqué la légitimité scientifique de ce concept.
- 2) Les concepts scientifiques sont des concepts objectifs.
- 3) La seule façon pour un concept scientifique d'être objectif est d'être indépendant des valeurs (le concept doit se baser uniquement sur les faits de la nature et exclure les valeurs sociales et culturelles).
- 4) La notion de dysfonction est indépendante des valeurs (objective).
- 5) Si le trouble mental est défini par la notion de dysfonction, alors le concept de trouble mental est scientifique
- 6) Le trouble mental est (ou doit être) défini par la notion de dysfonction.

(C)Alors, le concept de trouble mental est un concept scientifique et la psychiatrie peut répondre aux critiques qui ont attaqué son statut de science¹.

De manière générale, les critiques qui ont visé la NFO ont pris pour cible la prémisse 4. En effet, comme exposé dans les deux chapitres précédents, une grande majorité des critiques de la NFO ont soutenu que le concept de dysfonction n'était pas indépendant des valeurs sociales et culturelles, invalidant la prémisse 4. Bien que je sois d'accord avec cette ligne de critique, dans ce chapitre, je vais prendre pour cible les prémisses 2 et 3. Je souhaite défendre que les prémisses 2 et 3 s'ancrent dans ce qu'on a appelé l'« idéal de la science indépendante des valeurs ». Cet idéal est (du moins en partie) ce

¹ On pourrait objecter qu'un autre argument semble motiver les défenseurs de l'OFN. Il s'agit de l'idée selon laquelle le concept de dysfonction est bel et bien indépendant des valeurs. Or, sans l'argument tout juste présenté selon lequel un concept indépendant des valeurs est supérieur à un concept influencé par les valeurs pour fonder la distinction entre le normal et le pathologique, il n'y a pas de raison de prioriser un concept indépendant des valeurs par rapport à un concept influencé par des valeurs. Le fait que le concept de dysfonction soit indépendant des valeurs n'est pas en lui seul un argument.

pour quoi la NFO a cherché à élaborer une définition neutre de la santé mentale en utilisant la notion de dysfonction. Or cet idéal n'est pas sans problème, ce que je montrerai dans les prochaines sections. Si nous rejetons cet idéal, c'est du moins ce que je veux défendre, il devient alors possible d'élaborer un nouvel idéal auquel la psychiatrie pourrait peut-être se conformer.

4.2 La conception traditionnelle de la science : l'idéal de la science indépendante des valeurs et l'objectivité comme neutralité

Il s'agit désormais de comprendre en quoi les prémisses 2 et 3 s'ancrent dans l'idéal de la science indépendante des valeurs et les raisons pour lesquelles cet idéal devrait être rejeté en faveur de l'idéal d'objectivité sociale. Dans cette section, j'examinerai donc la thèse selon laquelle la science est et doit être détachée des valeurs, et le concept d'objectivité qui est associé à cette thèse. Dans nos sociétés contemporaines, la science occupe un rôle croissant. L'autorité et la valeur sont le plus souvent attribuées à la science parce qu'elle a l'ambition de produire un savoir sur le monde tel qu'il est, et non pas sur ce que nous pensons qu'il devrait être. La science est donc perçue comme une activité de production de faits sur le monde, par contraste avec, entre autres, l'art ou la littérature qui relèveraient davantage du domaine des jugements normatifs d'ordre éthique ou esthétique. Il est communément admis que les scientifiques formulent des hypothèses visant à décrire et à expliquer la réalité, qui auront à être corroborées ou infirmées par des données empiriques. Ce procédé est généralement perçu comme *objectif*, parce qu'en se consacrant uniquement aux faits empiriques, les scientifiques excluent les valeurs sociales et politiques qui pourraient venir fausser ou altérer les connaissances scientifiques (p. ex. : Haely, 2008, p. 6-7). Dans ce contexte, la science

est conçue comme étant objective, au sens où elle décrit les phénomènes du monde avec justesse. En raison de l'autorité croissante associée à la science, il n'est pas surprenant que plusieurs philosophes des sciences se soient intéressés aux critères qui permettent de déterminer ce qui est une « bonne science » et à l'objectivité scientifique (p. ex. : Daston et Galison, 2007; Douglas, 2004; Lloyd 1995).

La notion classique d'objectivité est intimement liée à la conception traditionnelle de la science comme « idéal de la science indépendante des valeurs ». Cette signification de l'objectivité et la conception de la neutralité de la science ont été théorisées par plusieurs auteurs (p. ex. : Biddle, 2013; Dorato, 2004; Ruphy, 2006;). Je reprends ici l'analyse de Heather Douglas puisqu'elle met l'accent sur la relation entre les valeurs et l'objectivité. Douglas décrit la conception de l'idéal de la science indépendante des valeurs ainsi : « the value-free ideal is more accurately the “internal scientific values only when performing scientific reasoning” ideal » (Douglas, 2009, p. 45). Selon cette interprétation, l'idéal de la science indépendante des valeurs peut être compris selon trois caractéristiques :

- 1) Les valeurs sociales, culturelles ou politiques (appelées également valeurs non épistémiques) sont acceptées dans les phases externes de la pratique scientifique. Autrement dit, les valeurs non épistémiques peuvent jouer un rôle dans la détermination de l'orientation des projets de recherche, les contraintes éthiques imposées sur les choix méthodologiques, ou comme *impetus* personnel motivant les chercheurs à poursuivre leur projet de recherche.
- 2) Les valeurs non épistémiques sont à proscrire dans les phases internes de la science. Cela signifie que les valeurs non épistémiques devraient être évitées dans les processus de raisonnement et de justification qui mènent à l'acceptation ou au rejet de théories ou d'hypothèses scientifiques.
- 3) Les valeurs épistémiques sont quant à elles acceptées dans les phases internes de la science. En d'autres mots, les valeurs relatives à la production de

connaissances scientifiques sont permises dans ces phases. La simplicité, le pouvoir explicatif, le pouvoir prédictif, l'étendue, la vérité, etc. peuvent intervenir dans les phases de raisonnement et de justification.

Ainsi, selon l'idéal de la science indépendante des valeurs, les scientifiques peuvent faire intervenir des valeurs non épistémiques dans les phases externes de la science, et des valeurs épistémiques dans les phases internes de la science. Cependant, ils ou elles devraient tenter, autant que faire se peut, d'éliminer l'influence des valeurs non épistémiques lorsque des théories ou des hypothèses sont testées ou évaluées à la lumière d'éléments de preuve.

Au sein de la vision traditionnelle de la science se dessine la conception classique de l'objectivité, qui peut être reconstruite selon trois éléments principaux (Crasnow, 2013) :

- 1) Le premier élément est parfois appelé « la conception absolue » ([the absolute conception], Williams, 1985) ou « la vue de nulle part » ([*the view from nowhere*], Nagel, 1986). Il s'agit de l'idée selon laquelle l'objectivité serait « sans perspective » (*aperspectival*), et que la science décrirait par là les objets du monde tels qu'ils sont, indépendamment des positions sociales, politiques, historiques ou culturelles dans lesquelles se trouvent les scientifiques, voire indépendamment de l'esprit humain.
- 2) Le deuxième élément est l'absence de valeurs non épistémiques. Comme décrit par l'idéal de la science indépendante des valeurs, seules les valeurs épistémiques peuvent être impliquées dans la justification épistémique. Dans la conception de l'idéal de la science indépendante des valeurs, l'intrusion de valeurs non épistémiques est une menace à l'objectivité scientifique étant donné que le but de la science est de représenter le monde tel qu'il est (voir Potter, 2006, p. 76-77 pour une discussion de ce point spécifique).

- 3) Le troisième élément est l'affirmation qu'il faut procéder, autant que possible, à l'élimination des biais personnels. Selon cet élément, les résultats scientifiques ne doivent pas être une simple projection des préférences individuelles des scientifiques. La science doit alors faire la promotion de procédures impartiales, dont le but est de prévenir l'influence de tels biais.

Ces trois éléments peuvent être pensés comme étant conjointement constitutifs de l'objectivité, ou lui être associés de manière individuelle.

4.3 La philosophie féministe des sciences et le concept d'objectivité

Cette conception de la science indépendante des valeurs et de l'objectivité comme neutralité ou impartialité a été abondamment critiquée dans les écrits en philosophie des sciences (p. ex. : Dupré, 2007; Feyerabend, 1962; Kuhn, 1962/1970). Un courant important qui a participé à cette remise en question est la philosophie féministe des sciences. Cette perspective a porté un dur coup à cette conception en s'intéressant à l'apport des femmes et des autres groupes marginalisés à la production de connaissances scientifiques. Selon la philosophe féministe des sciences Kristen Intemann, la philosophie féministe des sciences a généralement poursuivi trois objectifs :

- 1) Identifier et prévenir les biais de genres et les autres formes de biais problématiques dans le raisonnement scientifique;

- 2) Identifier et éliminer les obstacles formels et informels qui affectent la participation des femmes et des autres groupes marginalisés au sein de la science;
- 3) S'assurer que la connaissance, les interventions et les nouvelles technologies générées par l'investigation scientifique puissent bénéficier à tout le monde en société (par opposition à seulement quelques-uns, généralement les groupes en position de privilège social). (Intemann, 2017, p. 130-132)

Alors que les deuxième et troisième objectifs font généralement consensus et participent à un projet plus général de justice sociale, le premier objectif est plus litigieux. La philosophie féministe des sciences a en effet permis de mettre au jour l'influence de biais sexiste ou raciste au sein de la science, mais la manière de faire face à cette présence diffère en fonction des approches.

Pour mieux comprendre cette absence de consensus, je vais d'abord exposer certains exemples pour ensuite montrer comment ils peuvent être traités différemment en fonction des approches adoptées. Un exemple paradigmatique de l'influence induite de valeurs sexistes est certainement la théorie de l'infériorité cognitive naturelle des femmes, communément acceptée au 19^e siècle et au début du 20^e siècle. Cette théorie était perçue comme scientifique, car basée sur les mesures de certaines sections du crâne, ainsi que sur certaines interprétations de la théorie de l'évolution. Elle ne trouve aujourd'hui plus aucun soutien empirique (voir Belkhir, 1994; Shields, 1975, 1982, 2007). Un autre exemple bien connu en philosophie féministe des sciences est celui des modèles reproductifs qui, jusqu'en 1980, promouvaient la « saga du sperme » (*sperm saga*). Selon ces théories, le processus de fécondation humaine était faussement conçu comme le résultat de l'action des spermatozoïdes « actifs », décrits comme des « guerriers » et des « vaillants héros » par rapport à des ovules « passifs », voire « endormis ». Une fois la présence de biais sexistes mise en évidence, il a été possible d'observer un processus beaucoup plus interactif entre le spermatozoïde et l'ovule et

une reconstruction du langage servant à décrire ce phénomène (voir Haely, 2008, p. 69-71 ; Martin, 1991 pour une discussion).

Bien qu'il paraisse évident que dans ces deux exemples, les valeurs sexistes aient faussé la description des phénomènes à l'étude, la manière de surmonter ces biais crée polémique en philosophie des sciences et même parmi les philosophes féministes des sciences (voir Ruphy, 2015 pour une discussion). Une première option consiste à qualifier les épisodes scientifiques mentionnés plus haut de « mauvaise science » en ce sens qu'elle ne respecte pas l'idéal de neutralité de la conception traditionnelle de la science. L'intrusion de valeurs non épistémiques, ici sexistes, aurait contribué à fausser les résultats de recherche. En ce sens, l'élimination de valeurs non épistémiques constituerait la solution à leur influence nocive. Cependant, une autre option, plus polémique, a également émergé en philosophie féministe des sciences. Dans le but de minimiser l'influence des attitudes racistes ou sexistes en science, la majorité des philosophes féministes des sciences ont défendu l'idée que les valeurs féministes pouvaient être bénéfiques à la science, par contraste avec la thèse selon laquelle pour éviter des biais racistes ou sexistes, toutes formes de valeurs non épistémiques devraient être éliminées. Un exemple classique de recherches fructueuses menées dans une perspective féministe est le travail en sciences sociales de Stewart et coll. (1997) sur le divorce, analysé ensuite par Anderson (2004; voir Crasnow *et al.*, 2015 pour une discussion). Très brièvement, voyons comment Anderson conçoit les avantages de l'approche féministe dans ce cas. L'auteure contraste deux approches sur le divorce, l'une plus classique qui valorise le modèle traditionnel familial dans lequel l'homme et la femme sont mariés toute leur vie (p. ex. : Wallerstein et Kelly, 1980), et l'autre ancrée dans un cadre féministe (Abigail *et al.*, 1997). L'approche traditionnelle conçoit le divorce comme une expérience de traumatisme et de tristesse pour les membres de la famille, alors que Stewart et ses collègues conçoivent le divorce comme une expérience d'adaptation continue qui peut être vécue positivement ou négativement. Anderson défend que bien que les deux approches fassent appel à des valeurs sur ce

qu'est le divorce, l'analyse féministe de Stewart et coll. est meilleure que celle de l'approche traditionnelle sur le plan empirique. Par exemple, les données collectées par Stewart et coll. étaient plus complètes et représentatives que celles de l'approche traditionnelle. En effet, parce que l'approche traditionnelle a assumé que le divorce était négatif à la base, la collecte de données a été influencée négativement (p. ex.: les expériences positives liées au divorce avaient tendance à être laissées de côté parce que considérées comme non pertinentes, la possibilité que le divorce soit vécu différemment par les membres d'une même famille n'avait pas été considérée, les mesures prises consistaient généralement en une évaluation des perturbations psychologiques entraînées par le divorce, etc.). La perspective féministe n'avait quant à elle pas assumé que les effets du divorce seraient négatifs, ce qui a permis de parvenir à une vision plus nuancée et plus représentative des effets du divorce sur les différents membres de la famille (p. ex. : 70 % des femmes interviewées pensaient que leur personnalité s'était améliorée après le divorce, plusieurs femmes pensaient que même si leur situation financière s'était détériorée avec le divorce, elles appréciaient davantage leur autonomie financière, etc.). Avec le cas des études féministes sur le divorce, Anderson démontre qu'une « bonne science » peut en ce sens être biaisée en faveur de valeurs féministes (voir aussi Haraway, 1988; Harding; 1986; Longino, 1990, 2002). Cette thèse ambitieuse de certaines philosophes féministes des sciences a généré une levée de boucliers de la part d'un pan plus conservateur de la philosophie des sciences, motivé par la croyance selon laquelle l'ouverture à des valeurs non épistémiques viendrait limiter, voire annihiler, l'objectivité et par là l'autorité de la science (p. ex. : Gross et Levitt, 1994; Haack, 1993, 2003; Richards, 1995). Selon ce courant, en renonçant à la neutralité de la science, le critère pour confirmer une hypothèse deviendrait la conformation à des valeurs sociales et politiques plutôt qu'à des éléments de preuve empiriques (je discute plus en détail ce problème dans le chapitre 5, voir la section 5.4 et 5.5). Dans ce contexte, le projet de la philosophie féministe des sciences est double (Crasnow, 2013) : le premier volet, critique, consiste à montrer que le caractère désirable et idéal du concept d'objectivité comme

indépendance des valeurs est infondé, voire nuisible. Le second volet, plus positif, consiste en une démonstration de la possibilité d'une science objective qui serait en même temps influencée par des valeurs non épistémiques.

Ainsi, la philosophie féministe des sciences a permis d'ouvrir un espace pour penser une conception de l'objectivité qui admet et rend possible la présence de valeurs non épistémiques au sein de la science comme une alternative à la vision traditionnelle. Dans ce qui suit, je vais présenter les différentes critiques de l'idéal et une conception de l'objectivité issues de la philosophie féministe des sciences : la conception de l'objectivité sociale (OS) d'Helen Longino. Par la suite, j'argumenterai le fait que ce type de théorie permet de penser le rôle et l'influence des valeurs sociales et culturelles dans la définition du trouble mental.

4.4 Les différentes critiques de l'idéal de la science indépendante des valeurs

Dans la section précédente, je me suis intéressée à la manière dont la philosophie féministe des sciences a joué un rôle majeur dans la remise en question de l'idéal de la science indépendante des valeurs en permettant de penser différemment les concepts traditionnels de science et d'objectivité. Ces préoccupations par rapport à l'idéal se retrouvent désormais dans plusieurs champs de la philosophie des sciences. Dans cette section, je vais présenter un résumé des différentes critiques adressées à l'idéal de la science indépendante des valeurs. Cet idéal a été critiqué sous plusieurs angles. Pour faciliter la discussion, je vais m'appuyer sur la classification de ces différentes critiques offerte par Betz (2013). Celle-ci découpe les critiques de l'idéal en deux catégories : 1) la critique sémantique (Dupré, 2007; Putnam, 2002) et 2) la critique méthodologique (Douglas, 2007, 2009; Longino, 1990, 2002)². Je présenterai ces deux critiques dans les sections 4.4.1, 4.4.2 et 4.5.

² Betz propose en note de bas de page (note 1) qu'un autre type de critique pourrait être identifié : la critique empirique, trouvée chez Cranor (1990, p. 139) ou Elliott (2011, p. 55-80). La thèse soutenue par cette approche est la suivante : l'idéal de la science indépendante des valeurs peut entraîner des conséquences préjudiciables. Cela s'expliquerait par le fait que les avis scientifiques pourraient demeurer sans résultat ou ne pas être écoutés si les valeurs sociales et culturelles ne sont pas intégrées, ou si les avis sont trop nuancés. Or comme le note Betz, davantage de soutien empirique devrait être fourni pour que cette critique puisse trouver sa justification.

4.4.1 La critique sémantique

La critique sémantique est basée sur l'idée de « concepts épais » (*thick concepts*) développée par Williams (1985; voir Roberts 2013 pour une discussion). Selon cette théorie, dans la signification de certains concepts, il y aurait une imbrication de faits et de valeurs impossible à défaire. Autrement dit, la signification de certains concepts serait intrinsèquement constituée de composantes factuelles et normatives, et il serait impossible de réduire cette signification à l'une ou l'autre de ces composantes. Les exemples que discute Williams (1985, p. 140) sont de l'ordre de l'éthique (p. ex. : « brutalité », « gratitude », « mensonge », « lâcheté »). Or cette conception a été étendue aux concepts scientifiques. Par exemple, pour Dupré (2007), bien que certains énoncés en science soient purement factuels (p. ex. : « les électrons ont une charge négative », p. 28), d'autres énoncés semblent difficilement entrer dans une seule catégorie, « factuelle » ou « évaluative », puisqu'ils semblent relever à parts égales des deux catégories (p. ex. : « Les États-Unis sont un pays violent. », p. 29). Dupré explique que cette difficulté à qualifier ces énoncés comme « évaluatifs » ou « factuels » serait liée à la signification de certains concepts scientifiques : leur signification serait dépendante de nos intérêts humains. Cette dépendance impliquerait que ces concepts scientifiques, bien que relevant de faits empiriques, soient également influencés par les valeurs sociales et culturelles. Certains concepts scientifiques référerait donc à la fois à des critères descriptifs et à une signification normative. Pour Dupré, cette imbrication de faits et de valeurs au sein de concepts scientifiques remettrait en question la distinction classique que nous traçons entre faits et valeurs, et remettrait par là en question l'idéal de la science indépendante des valeurs.

En guise d'illustration, Dupré discute des concepts tels que le « viol » en psychologie évolutionniste et l'« optimalité » en sciences économiques. Pour Dupré, ces concepts

ne peuvent pas être opérationnalisés par des composantes purement descriptives sans perdre leur signification originale. Dans le cas du viol, le concept est étudié en psychologie évolutionniste au regard des avantages évolutifs qu'il présente pour les individus mâles possédant peu de traits avantageux en termes de sélection sexuelle. Dans un environnement primitif, le cadre évolutionniste pose que les femmes avaient tendance à choisir des partenaires sexuels capables de leur donner les ressources nécessaires pour assurer leur subsistance et celle de leur progéniture. L'hypothèse donnée pour expliquer l'émergence du viol est généralement que les hommes possédant peu de ressources à offrir à leur potentielle partenaire de reproduction avaient comme option alternative la violence sexuelle, assurant ainsi une transmission des gènes. Le comportement de violence sexuelle aurait donc pu être sélectionné dans un environnement évolutif passé et se transmettre jusqu'à aujourd'hui en raison de l'avantage évolutif qu'il conférait alors aux hommes violents. Les hommes posséderaient donc un « module psychologique du viol », activé lorsque leurs possibilités d'attirer les faveurs sexuelles des femmes seraient trop basses (Dupré, 2007, p. 33). Selon les penseurs de la psychologie évolutionniste qui étudie le viol, ces résultats de recherche représentent uniquement des faits empiriques.

Dupré conteste cette opposition tranchée entre faits et valeurs. Selon lui, les explications évolutionnistes éludent complètement deux aspects importants du concept de viol : les présuppositions morales associées aux violences sexuelles ainsi que les causes structurelles et sociales qui peuvent aider à comprendre le phénomène :

Those who have thought seriously about contemporary sexual violence as opposed to the hypothetical reproductive strategies of imagined ancestors have observed that rape is not exclusively, or even mainly, a crime of resourceless reproductive predators lurking in dark alleyways but has much more to do with misogyny, and more to do with violence than sex, let alone reproduction. Its causes appear, therefore, to be at the level of ideology rather than economics (Dupré, 2007, p. 33).

Ainsi, la perspective évolutionniste ne peut pas rendre compte du viol compris comme un concept moralement condamnable (brimant les droits des victimes), ni comme un phénomène complexe qui rend manifestes les relations de pouvoir et les préjugés (tels que le patriarcat menant à la misogynie). Ces éléments sont fondamentaux pour comprendre ce que devrait être le viol (Dupré, 2007, p. 37). En effet, le concept de viol réfère davantage au fait de ne pas obtenir le *consentement* d'une personne avant l'acte sexuel. Selon Dupré, les conclusions auxquelles aboutissent les psychologues évolutionnistes sont trop distantes de ce qu'est réellement le viol parce que le concept de viol a été vidé de sa signification normative. Notre concept général de viol implique l'immoralité et la relation de pouvoir inégale entre les hommes et les femmes, et le non-respect du consentement. La psychologie évolutionniste, en offrant une explication purement factuelle de l'émergence du viol qui repose sur une analyse coûts/bénéfices en termes de *fitness*, ne rend pas compte de ce que nous désignons par le mot « viol ». Elle explique seulement un aspect du viol en insistant sur l'avantage évolutif engendré par ce type de violence sexuelle. En cela, la psychologie évolutionniste propose une explication restreinte, voire inadéquate, du viol, qui ne cadre pas avec la signification du concept. Le cas des études en psychologie évolutionniste sur le viol témoigne du fait qu'il n'est pas souhaitable, voire possible, d'éliminer les composantes normatives des concepts à l'étude puisque l'investigation scientifique risque de perdre de vue ce qu'elle cherchait réellement à expliquer. À l'aide d'autres exemples similaires à celui du viol, mais tirés du domaine de l'économie, Dupré en vient à la conclusion que certains concepts sont intrinsèquement composés de faits et de normes sociales et culturelles. Cette interconnexion est ce pour quoi la science ne peut pas être dite indépendante des valeurs dans son ensemble. Cette interconnexion est également ce pour quoi la science ne devrait pas nécessairement chercher cette indépendance, puisqu'elle risque d'aboutir à des conclusions trop partielles et éloignées de nos intérêts.

4.4.2 Heather Douglas et la critique méthodologique

Bien que la critique sémantique soit intéressante, c'est la critique méthodologique qui a été la plus influente en philosophie des sciences. Il existe deux types de critiques méthodologiques : l'argument du risque inductif, aussi appelé « argument de l'erreur » (Douglas, 2007, 2009; voir aussi Biddle et Kukla, 2017 pour une critique similaire et Brown, 2013 pour l'appellation) et l'argument de la sous-détermination des théories par les éléments de preuves, aussi appelé « argument de l'écart » (Longino, 1990, 2002; voir Brown, 2013 pour l'appellation). Dans cette section, je vais présenter l'argument du risque inductif tel que formulé par Douglas. Dans la section 4.5, je vais me pencher plus avant sur l'argument de la sous-détermination des théories par les éléments de preuve. Cet argument est fondamental dans la théorie de Longino et me permettra de montrer comment sa théorie de l'objectivité sociale y est rattachée. C'est pourquoi je lui consacre une section entière et indépendante.

Douglas (2000, 2003, 2007, 2008, 2009) s'inscrit dans une approche qui vise à questionner la relation entre la science et les politiques publiques. Elle défend que l'idéal de la science indépendante des valeurs est fondé sur le faux présupposé de l'isolation (réel ou prescrit) de la communauté scientifique par rapport au reste de la société. En raison du lien établi entre science et société dans les sociétés occidentales contemporaines, ce présupposé devrait être rejeté. Selon Douglas, les scientifiques doivent avoir recours à des jugements de valeur lorsque de l'incertitude subsiste quant à l'acceptation ou au rejet d'une hypothèse ou d'une théorie. Les scientifiques doivent prendre des décisions par rapport à la qualité et à la quantité d'éléments de preuve requis pour accepter une hypothèse dans des cas d'incertitude. Parce qu'ils ont une meilleure expertise quant aux conséquences d'une erreur scientifique, ce sont eux qui devraient intégrer à leur raisonnement des valeurs non épistémiques leur permettant

d'évaluer une théorie. La science ne devrait pas chercher à atteindre l'idéal de l'indépendance des valeurs.

La critique de Douglas est appelée « argument du risque inductif » (*inductive risk argument*). Dans plusieurs circonstances, les scientifiques font face à de l'incertitude quant à l'acceptation ou au rejet d'une hypothèse scientifique. Cette incertitude est liée à certains risques associés aux conséquences de l'acceptation ou du rejet d'une hypothèse scientifique dans un contexte d'incertitude. Pour mieux comprendre les devoirs moraux des scientifiques, Douglas trace une analogie entre la responsabilité morale que nous attribuons à tout un chacun, et celle que nous devrions attribuer aux scientifiques. Selon Douglas, nous attribuons une responsabilité morale à un individu pour ses choix dans des situations d'incertitude. Nous attendons d'un individu qu'il soupèse les risques entraînés par ses choix lorsque les conséquences de ses choix ne peuvent pas être connues absolument. Par exemple, si un individu trouve une valise abandonnée dans un métro, il fait face à une situation d'incertitude : la valise contient-elle une bombe, prête à exploser? L'individu n'est pas en mesure de le savoir avec exactitude. Il doit donc évaluer les risques associés au fait de croire que la valise contient une bombe et les risques associés au fait de croire que la valise a simplement été oubliée par un usager du métro. Si la valise contient une bombe, l'individu devrait appeler la police. Si la valise ne contient pas de bombe, l'individu devrait la laisser à sa place. Or étant donné que l'individu ne sait pas ce que contient la valise, il doit évaluer les risques d'erreur associés au fait de penser que la valise contienne une bombe ou non. Si la valise contient une bombe, mais que l'individu la laisse à sa place, des individus mourront, et un ensemble de biens sociaux matériels seront détruits. Si la valise ne contient pas de bombe, mais que l'individu appelle tout de même la police, alors les personnes faisant usage du transport en commun seront retardées sans raison par l'intervention de la police et seront ralenties dans leur déplacement. Dans cette situation, le risque d'erreur lié au fait de ne pas appeler la police est plus grand que le risque lié à l'intervention de la police. Une fois les risques

soupesés, l'individu aurait donc la responsabilité morale d'appeler la police (Douglas, 2009, p. 68-71).

De la même façon, les scientifiques sont placés dans des situations d'incertitude ou de risque inductif. Biddle a reformulé l'argument de Douglas comme suit :

Inductive risk is the risk of wrongly accepting or rejecting a hypothesis H, given a body of evidence E that is taken to support H. Given that scientific evidence cannot provide proof that a hypothesis is true, there is always some inductive risk of accepting H (given E) when H is false or rejecting H when H is true (Biddle, 2016, p. 193).

Dans ce contexte, les scientifiques auraient la responsabilité morale de procéder de manière similaire, c'est-à-dire d'évaluer l'incertitude en jeu ainsi que les risques d'erreur associés au fait d'accepter ou de rejeter une théorie. Les situations d'incertitude dans lesquelles sont placés les scientifiques ressemblent à l'analogie présentée par Douglas. Se basant sur les travaux de Churchman (1948), de Rudner (1953) et d'Hempel (1965), Douglas défend l'idée que les scientifiques ont à prendre des décisions par rapport à la quantité et à la qualité du soutien empirique requis pour accepter ou rejeter une hypothèse. Ces décisions peuvent alors être motivées par des valeurs non épistémiques, qui permettent de déterminer le degré de soutien empirique requis pour accepter ou rejeter une hypothèse. Dans ce type de cas, les scientifiques devront prendre en considération les conséquences entraînées par une erreur (adopter une hypothèse fautive ou rejeter une hypothèse vraie). La thèse avancée par Douglas est soutenue par de multiples études de cas où les composantes empiriques seules ne permettent pas de confirmer ou de rejeter une hypothèse. Les possibles conséquences néfastes associées à l'adoption ou au rejet erronés d'une hypothèse sont ce qui vient compléter le raisonnement scientifique. Dans ces circonstances, les scientifiques doivent intégrer des valeurs éthiques dans leur raisonnement en fonction des conséquences associées à une erreur.

Un exemple parlant utilisé par Douglas est celui de recherches portant sur les effets cancérigènes de la dioxine, un composé chimique présent dans certains pesticides, sur le rat. Les écrits scientifiques sur le sujet étaient controversés par rapport au lien entre une exposition faible à la dioxine et le développement de certaines formes de cancer. La controverse s'expliquant par le fait qu'il était difficile pour les scientifiques d'interpréter les diapositives montées à partir de foie de rat, en ce que la distinction entre une tumeur bénigne et une tumeur maligne n'était pas évidente au regard de ces seuls éléments de preuve (voir Biddle, 2016; Biddle et Kukla, 2017; Kukla, 2019 pour une discussion sur cet exemple précis et la notion de risque inductif et ce qu'ils appellent la « production d'éléments de preuve »). Étant donné que l'hypothèse « La dioxine cause le cancer » peut avoir des effets majeurs sur la santé publique (en ce que cela déterminera si la dioxine sera étiquetée comme cancérigène et si sa vente sera limitée), il est possible de demander un soutien empirique moins ambigu (Douglas, 2000, p. 569-572). Le risque d'erreur associé à cette hypothèse étant très sérieux, il est de la responsabilité des scientifiques de prendre en considération les enjeux moraux reliés à l'acceptation ou au rejet de cette hypothèse (voir aussi Douglas, 2009)³.

Certains pourraient vouloir défendre que les scientifiques, de par le fait qu'ils ont aussi la responsabilité de viser à atteindre les buts de la science (améliorer les prédictions et les explications à propos du monde, ce que Douglas appelle les « responsabilités de rôle [*role responsibilities*]), devraient être exemptés de la responsabilité morale d'évaluer

³ Un point de précision s'impose ici : Douglas ne dit pas que les valeurs devraient remplacer les éléments de preuve dans l'acceptation ou le rejet des hypothèses. Elle distingue deux types de rôles que peuvent avoir les valeurs sociales et culturelles : un rôle direct et un rôle indirect. Le rôle direct des valeurs non épistémiques en science intervient lorsque ces valeurs sont des raisons en elles-mêmes pour accepter ou rejeter une hypothèse (as « reasons in themselves to accept a claim »). Les valeurs jouent un rôle indirect lorsqu'elles sont des raisons pour déterminer le degré de soutien empirique requis pour accepter ou rejeter une hypothèse. Selon Douglas, le rôle légitime des valeurs en science est indirect. (voir Douglas, 2009, p. 94-112 pour une discussion complète; Brown, 2013 pour un commentaire; Elliott, 2011, 2013 et Steel et Whyte, 2012 pour une critique.)

les potentielles conséquences d'erreur liées à leur choix. Or selon Douglas, il serait néfaste d'exempter les scientifiques d'une telle responsabilité étant donné qu'ils sont ceux ayant la plus grande expertise des incertitudes et des risques d'erreurs reliés à leur investigation scientifique (Douglas, 2009, p. 71-74). Douglas écrit :

This is the weighing of uncertainty around an empirical claim against the ethical consequences of error. In making crucial empirical claims about the likelihood of error, scientists should consider the consequences of error, for such consideration is what requires different burdens of proof in different contexts (Douglas, 2009, p. 78-79).

Les jugements de valeur deviennent donc une part importante du raisonnement scientifique lorsque les scientifiques évaluent le degré d'incertitude et les risques d'erreurs lui étant associés. Pour Douglas, en fonction de l'argument du risque inductif, l'idéal de la science indépendante des valeurs doit donc être rejeté au profit d'une intégration « contrôlée » des valeurs sociales et culturelles dans le raisonnement scientifique.

À ce stade, il est raisonnable de se demander si les critiques de l'idéal de la science indépendante des valeurs ne concernent pas seulement certains contextes d'investigation scientifique plus pratiques. Avant de clore cette section, il convient en effet de noter que la thèse que je défends n'est pas liée au succès des critiques de l'idéal de la science indépendante des valeurs dans tous les domaines d'investigation scientifique. Par exemple, des chercheurs défendent que certaines branches de la science, comme la physique ou la chimie, puissent être dites indépendantes des valeurs, et ainsi répondre à l'idéal. Par exemple, le représentant de la critique sémantique, Dupré (2007), écrit :

What I want to say about physics is that if most or all of physics is value free, it is not because physics is science but because most of physics simply doesn't

matter to us. Whether electrons have a positive or a negative charge and whether there is a black hole in the middle of our galaxy are questions of absolutely no immediate importance to us. The only human interests they touch (and these they may indeed touch deeply) are cognitive ones, and so the only values that they implicate are cognitive values. The statement that electrons have negative charge is thus value free in a quite banal sense: It has no bearing on anything we care about. (Dupré, 2007, p. 32).

Ce que l'ensemble des critiques présentées et l'approche de la philosophie féministe des sciences montrent est qu'il y a de bonnes raisons de douter de l'idéal de la science indépendante des valeurs dans plusieurs contextes de recherches scientifiques, principalement celles ayant des implications sociales et culturelles ou relevant des intérêts humains. Ainsi, il devient raisonnable de douter des prémisses 2 et 3 de l'argument de la NFO, c'est-à-dire qu'il est possible de remettre en question l'idée que seuls les concepts purement indépendants des valeurs peuvent être dits « scientifiques » et « objectifs ». Bien qu'il soit possible que certains domaines de recherche puissent atteindre cet idéal, dans la section 4.7, je montrerai en quoi les prémisses 2 et 3 peuvent nous aiguiller sur les risques que peuvent entraîner l'intrusion de valeurs en science pour les populations vulnérables. J'y explorerai la manière dont cette intrusion peut être problématique dans le cas de la psychiatrie et de la définition du trouble mental.

4.5 La critique méthodologique et l'empirisme contextualiste critique d'Helen Longino

Dans la section précédente, j'ai présenté les différentes critiques ayant ciblé l'idéal de la science indépendante des valeurs. Comme mentionné plus tôt, l'attaque contre cet

idéal ayant eu le plus d'influence est la critique dite « méthodologique ». Le volet le plus connu de cette critique est sans contredit l'œuvre d'Helen Longino, et sa théorie de l'empirisme contextualiste critique (ECC). L'ECC se présente comme une théorie alternative de la justification épistémique, conforme à la présence de valeurs en science. La théorie de Longino a connu une forte notoriété parce qu'elle propose à la fois une critique détaillée de l'idéal de la science indépendante des valeurs et une solution originale au problème de la sous-détermination des théories par les éléments de preuves : l'objectivité sociale. En cela, elle répond au double projet de la philosophie féministe des sciences (voir la section 4.3). En effet, l'ECC est basé sur la thèse de la sous-détermination des théories scientifiques par les éléments de preuve défendue par Duhem et Quine. Longino reprend cette thèse pour en faire une critique de l'idéal et en tire les conséquences en élaborant un nouvel idéal pour la science. Dans ce qui suit, je présenterai la thèse de la sous-détermination et montrerai comment cette thèse conduit à la critique méthodologique de la science indépendante des valeurs. Dans la section 3.6, je vais montrer comment l'idéal de l'objectivité sociale vient pallier les nouveaux problèmes entraînés par la thèse de la sous-détermination.

La thèse de la sous-détermination a d'abord été formulée par P. Duhem, puis reprise par W.V.O. Quine⁴. Pour bien comprendre cette thèse, il faut la situer dans son contexte. Duhem s'oppose à la représentation de la relation de confirmation entre les hypothèses et les théories scientifiques promue par l'empirisme logique. L'empirisme logique concerne la justification épistémique des théories scientifiques, systématisée de manière logique. Cette approche traite également de la confirmation des énoncés observationnels. L'empirisme logique présente le raisonnement scientifique sous la forme hypothético-déductive, qui consiste en l'articulation d'une hypothèse et en la

⁴ Cette thèse a pris plusieurs formes, voir Intemann 2005, Brown 2013 et Stanford 2017 pour une discussion. Ici, je reprends la forme classique défendue par Duhem et Quine, puisque c'est sur cette formulation que se base l'ECC de Longino. Cette formulation est parfois appelée l'« indétermination holiste » (*holist underdetermination*) (voir Turnbull, 2017).

spécification des conséquences observables relatives à cette hypothèse. Dans ce cadre, une hypothèse est confirmée si les conséquences correspondent à celles attendues par l'hypothèse. Pour l'empirisme logique, la méthode des philosophes est celle de la reconstruction rationnelle : l'organisation d'une théorie en un système déductif. L'évaluation d'une théorie se fait donc par des tests de cohérence et de conséquences observationnelles. Les éléments qui ne peuvent pas être liés formellement au contenu observationnel des théories sont éliminés. Selon Duhem, le principal problème que rencontre l'empirisme logique est le suivant : les hypothèses théoriques n'impliquent jamais à elles seules les conséquences observationnelles qui serviront à les tester; il y a toujours des présuppositions qui sont nécessaires pour établir la pertinence des conséquences observationnelles pour des hypothèses données. Autrement dit, une hypothèse ne peut pas être testée de manière isolée, puisque la formulation de cette hypothèse demande que l'on fasse un certain nombre de présuppositions théoriques ou méthodologiques.

Un exemple classique est celui de la mécanique céleste de Newton, et plus précisément de l'orbite d'Uranus et de la découverte de Neptune. La mécanique céleste de Newton ne parvenait pas à prédire l'orbite d'Uranus. Au lieu de rejeter la théorie élaborée par Newton au regard de sa prédiction incorrecte, les scientifiques de l'époque ont plutôt rejeté l'hypothèse selon laquelle le système solaire comprenait seulement sept planètes. Les astronomes de l'époque ont alors fait l'hypothèse de l'existence d'une huitième planète et ont calculé son emplacement à partir de la position de l'orbite d'Uranus. C'est de cette façon que les scientifiques ont découvert Neptune, en 1846. La même stratégie a ensuite été utilisée pour expliquer le périhélie dans l'orbite de Mercure en postulant l'existence d'une neuvième planète, « Vulcain », qui aurait été située entre Mercure et le soleil. Or cette stratégie échoue. Il faudra l'élaboration de la théorie de la relativité générale d'Einstein pour parvenir à expliquer le phénomène. Cet exemple est utilisé pour illustrer l'indétermination des théories par les éléments de preuve (voir Stanford, 2017 pour une discussion plus détaillée de l'exemple). La relation entre les

théories et les éléments de preuve n'est pas indépendante : elle dépend d'autres hypothèses et éléments théoriques. Ici, la théorie de la mécanique céleste de Newton dépendait d'autres éléments théoriques comme l'existence de sept planètes dans le système solaire. C'est cette hypothèse qui a été rejetée face à la prédiction incorrecte de la position de l'orbite d'Uranus, et non pas la théorie newtonienne. Ce type d'exemple servait à montrer que la relation de confirmation ou de réfutation (*disconfirmation*) ne peut pas être représentée selon le modèle hypothético-déductif. Cette affirmation a pour implication que si les scientifiques n'observent pas les conséquences attendues par les hypothèses, on ne peut pas savoir, d'un point de vue logique, si c'est l'hypothèse ou la présupposition qui est réfutée. Il existerait donc un écart logique entre les énoncés qui articulent les relations théoriques et les énoncés qui décrivent les données constituant les éléments de preuve de ces énoncés.

Tout comme ces deux auteurs, Longino défend l'idée que les données, à elles seules, ne peuvent pas départager des hypothèses/théories concurrentes. Cette idée est soutenue par le fait que les hypothèses/théories concernent des éléments « extra-empiriques » distincts de ce qui est décrit dans les énoncés concernant les données. Selon elle, la relation entre les hypothèses théoriques et les éléments de preuve est toujours médiée par des présuppositions d'arrière-plan, qui forment le « contexte » dans lequel les données acquièrent leur pertinence en tant que preuve (Longino, 1990, p. 44). Dans la perspective contextualiste adoptée par Longino, différents présupposés d'arrière-plan vont assigner une pertinence en termes d'éléments de preuve à un même ensemble de données et/ou d'observations. En guise d'illustration, Longino utilise deux hypothèses concurrentes d'anthropologie évolutionniste sur le rôle des hommes et des femmes dans l'évolution culturelle humaine, et plus spécifiquement sur l'émergence des outils : « l'homme chasseur » [man the hunter] et « la femme cueilleuse » [woman the gatherer] (Longino, 1990, p. 107-109; voir aussi Potter, 2006 pour une discussion plus détaillée). L'approche de l'homme chasseur stipule que le développement des outils chez l'humain est attribuable au changement de comportement des hommes, qui

se sont mis à chasser. L'approche de la femme cueilleuse pose quant à elle que les outils ont émergé en raison d'un changement dans l'environnement impliquant des pratiques de cueillettes différentes et des conditions de reproduction modifiée. Différents éléments de preuve sont disponibles pour émettre ces hypothèses : des fossiles d'os, de dents et de différents objets, des connaissances connexes sur le développement du cerveau, du langage et de la coopération, etc. Ces mêmes éléments sont utilisés par les deux approches pour parvenir à des hypothèses différentes : ils servent de preuves du développement d'armes de chasse dans l'hypothèse de l'homme cueilleur, alors qu'ils servent de preuve au développement d'outils de récolte et d'outils en pierre pour ramollir les fibres dures ou écraser les cosses de graines dans l'hypothèse de la femme cueilleuse. À eux seuls, les éléments de preuve ne permettent pas réellement de trancher entre les deux explications avancées. Longino écrit :

If female gathering behavior is taken to be the crucial behavioral adaptation, the stones are evidence that women began to develop stone tools in addition to the organic tools already in use for gathering and preparing edible vegetation. If male hunting behavior is taken to be the crucial adaptation, then the stones are evidence of male invention of tools for use in the hunting and preparation of animals. (Longino, 1990, p. 111)

C'est donc les présuppositions d'arrière-plan, ici le choix préalable du modèle de l'homme chasseur ou de la femme cueilleuse, qui permettent d'interpréter les éléments de preuve en faveur de l'une ou l'autre des approches. La seule consultation des éléments de preuve n'est pas suffisante.

Ici, une préoccupation légitime peut être soulevée lorsque l'on reconnaît la présence des présuppositions d'arrière-plan dans la justification épistémique est la possible – et plausible – influence des valeurs et des intérêts dans la production de connaissances scientifiques. Si la connexion de justification établie entre les théories et les éléments de preuve n'est pas seulement déterminée par un lien logique, alors les valeurs sociales

et culturelles peuvent jouer un rôle dans cette connexion (Longino, 1983, 1990, chapitres 5 et 6). En effet, les valeurs non épistémiques pourraient agir à titre de présuppositions d'arrière-plan ou pourraient motiver l'acceptation de présuppositions d'arrière-plan. La connaissance scientifique peut alors être dite influencée par les valeurs parce que les présuppositions d'arrière-plan ne sont pas composées uniquement de faits scientifiques et de valeurs épistémiques, mais également de valeurs non épistémiques.

Longino conçoit la présence des valeurs dans les phases internes de la science en introduisant une distinction entre valeurs *constitutives* et valeurs *contextuelles*, une distinction similaire à celle entre valeurs épistémiques et non épistémiques. Pour Longino, les valeurs constitutives sont les valeurs associées aux buts que nous attribuons à la science. Ces valeurs vont généralement fonctionner comme des règles ou des normes guidant les pratiques scientifiques et l'acceptation des résultats scientifiques. La vérité, la simplicité et la prédictivité sont des exemples de valeurs constitutives. Par contraste, les valeurs contextuelles réfèrent aux valeurs personnelles, sociales et culturelles qui sont propres au contexte dans lequel une investigation scientifique est menée (Longino 1983, p. 7-8, 1990, p. 4). Généralement, les valeurs constitutives, à l'image des valeurs épistémiques, ne sont pas jugées problématiques par les tenants de l'idéal de la science indépendante des valeurs. Ce sont plutôt les valeurs contextuelles qui sont perçues comme les ennemies d'une « bonne » science. Rappelons que dans l'idéal décrit par Douglas, il est traditionnellement défendu que les valeurs contextuelles (non épistémiques) devraient être éloignées de la science, du moins dans les phases internes de la science (Longino, 1990, p. 85). L'originalité de Longino réside dans son rejet de cette distinction traditionnelle entre valeurs contextuelles et valeurs constitutives et dans son intégration des valeurs contextuelles au sein des étapes internes de la science.

Les valeurs contextuelles interviennent de différentes façons dans les étapes internes de la science (Longino, 1983, p. 9-15, 1990, chapitre 5 et 6). L'une de ces façons est par le biais des présuppositions d'arrière-plan. Une valeur contextuelle pourrait jouer le rôle de présupposition d'arrière-plan ou pourrait motiver l'acceptation de présuppositions d'arrière-plan. Ces présuppositions d'arrière-plan, composées ou influencées par les valeurs contextuelles, sont ensuite utilisées pour faire des inférences dans des domaines de recherche spécifiques (Longino, 1990, p. 86). La connaissance scientifique peut alors être dite influencée par les valeurs parce que les présuppositions d'arrière-plan ne sont pas composées uniquement de faits scientifiques et de valeurs constitutives, mais également de valeurs contextuelles. Ces affirmations sont fondées sur plusieurs études de cas dont s'inspire Longino, dont celle de la différenciation sexuelle et l'endocrinologie (Longino, année, chapitres 6-8), que je vais maintenant discuter.

Dans les écrits scientifiques, il est admis qu'il existe une corrélation entre les hormones sexuelles et la différenciation sexuelle, et ce, dans trois domaines : les différences anatomiques et physiologiques, les différences comportementales et les différences cognitives (Longino, 1990, p. 112). Cette corrélation est interprétée comme la manifestation de l'effet des hormones sexuelles sur la différenciation sexuelle. L'objectif de Longino dans cette étude de cas est de mettre au jour les présupposés d'arrière-plan qui font partie du contexte social et culturel dans lequel ces recherches en endocrinologie sont menées. Dans ce qui suit, il s'agira de présenter la discussion de Longino qui porte sur les différences sexuelles et anatomiques liées à l'effet des hormones sexuelles, puisqu'il s'agit du cas où les recherches empiriques à ce sujet sont les plus solidement établies. Il s'agit donc d'un cas que l'on pourrait qualifier de « bonne science », dont les résultats sont fondés sur des éléments de preuve solides⁵.

⁵ Les deux autres cas discutés par Longino constituent des exemples où les résultats empiriques ne permettent pas de conclure à un effet direct des hormones sexuelles sur la différenciation comportementale ou cognitive. Il est donc moins clair que Longino considère ces deux cas comme de la

Comme le rapporte Longino, trois types d'éléments de preuve sont utilisés pour démontrer l'effet des hormones sexuelles sur la différenciation sexuelle anatomique, soit :

- 1) La corrélation entre l'anatomie mâle et femelle et un niveau plus ou moins haut d'hormones sexuelles (androgènes et œstrogènes) dans le corps;
- 2) La corrélation entre l'anatomie masculine et féminine « ambiguë » et des niveaux très hauts ou très bas d'hormones sexuelles;
- 3) Les études sur des animaux portant sur les effets des différents niveaux d'hormones et la différenciation sexuelle (Longino, 1990, p. 113).

Longino montre qu'au sein des études portant sur la différenciation sexuelle anatomique et physiologique en fonction de l'endocrinologie, la valeur (sociale) du dimorphisme sexuel (ou de l'essentialisme sexuel) joue un rôle clé dans la relation d'évidence que les chercheur.e.s formulent entre les données empiriques et les hypothèses. Cette valeur contextuelle ferait partie des présupposés d'arrière-plan globaux qui constitueraient le contexte de la recherche en endocrinologie. Autrement dit, selon Longino, ces études sont menées en fonction de la présupposition selon laquelle il existerait deux sexes biologiques, les sexes masculin et féminin, clairement différenciés. Ainsi, chacun des trois types d'éléments de preuve devient pertinent pour corroborer la différenciation sexuelle dimorphique seulement si le dimorphisme est déjà accepté en tant que valeur contextuelle (en tant que présupposé d'arrière-plan). Pour Longino, bien que cette valeur entre en jeu dans le processus de justification épistémique, il n'en demeure pas moins que les résultats scientifiques sont appuyés sur des données empiriques. Selon Longino, l'admission du rôle du dimorphisme comme présupposé n'invalide pas l'entièreté des recherches menées à ce sujet. Le danger réside

« bonne science », ce pour quoi je ne les discute pas ici.

plutôt dans le fait que si ces présomptions concernant le dimorphisme sexuel ne sont pas explicitées, elles peuvent masquer la possibilité d'un classement différent, par exemple un classement des sexes biologiques sur un continuum, sur lesquels les prototypes sexuels masculins et féminins peuvent représenter les extrêmes (Longino, 1990, p. 126-127; voir aussi Crasnow, 2001 pour ce type d'analyse). Autrement dit, dans ce cas d'étude, ce n'est pas le fait qu'une valeur non épistémique telle que le dimorphisme ait influencé la justification épistémique qui est problématique. Les deux classifications, celle du continuum des sexes biologiques et celle du dimorphisme, sont possibles au regard des données empiriques recueillies. Ce qui pose problème dans ces études est le présupposé du dimorphisme sexuel qui n'a pas été explicité comme présupposé, et qui est plutôt présenté comme un résultat empiriquement prouvé.

Par ce cas d'étude, il est possible de comprendre que c'est par le biais des valeurs non épistémiques présentes dans les présuppositions d'arrière-plan que la science s'éloigne de l'idéal de neutralité décrit plus tôt. Ce type de critique méthodologique ajoute donc du poids en faveur du rejet de l'idéal de la science indépendante des valeurs. Ce rejet permet de son côté l'élaboration d'une nouvelle conception de l'objectivité qui admet les valeurs non épistémiques au cœur même des phases internes de la science.

Avant de conclure cette section, il faut noter, suivant l'interprétation d'Intemann (2005), que les valeurs non épistémiques ne sont pas nécessairement présentes dans les présuppositions d'arrière-plan de *toutes* les hypothèses et théories scientifiques. En effet, Longino ne dit pas qu'il y a toujours des présuppositions d'arrière-plan qui contiennent des valeurs non épistémiques, mais plutôt qu'on ne peut pas savoir *a priori* si elles sont présentes ou non (p. ex. : Longino, 2002, p. 50). Dans cette optique, il est possible de supposer que ce n'est pas la science dans sa totalité qui est influencée par des valeurs non épistémiques, mais que *la possibilité d'influence* est à prendre au sérieux partout en science. La tâche de Longino est alors de développer un modèle qui

permette d'identifier les lieux où les valeurs non épistémiques entrent en jeu, et d'évaluer si leur rôle y est problématique.

4.6 Pour un nouvel idéal : l'objectivité sociale (OS)

Suite au rejet de l'idéal de la science indépendante des valeurs et à l'ECC, c'est plutôt l'idéal de l'objectivité sociale (OS), développée par Longino, qui selon elle devrait fournir à la science ces nouvelles balises. Au sein de ce modèle, les valeurs non épistémiques peuvent avoir un rôle légitime. En tenant pour acquise la présence plus que plausible des valeurs non épistémiques en science, c'est le dialogue critique entre les membres d'une communauté scientifique qui devrait permettre d'identifier et de critiquer l'influence des valeurs et des intérêts au sein des différentes communautés scientifiques. Cette conception de l'objectivité est caractérisée de *sociale*, car pour Longino, seule une communauté épistémique peut être dite objective. Elle est aussi qualifiée de « procédurale » dans les écrits parce que ce sont les procédures et méthodes de production de savoir scientifique mises en place par une communauté scientifique qui font de cette communauté épistémique une communauté objective (p. ex. : Hicks, 2011). Avec l'OS, Longino offre une conception de l'objectivité qui permet aux valeurs non épistémiques de jouer un rôle dans la connaissance scientifique, tout en ayant l'ambition de rejeter le relativisme radical. Au cœur de sa théorie se trouve la notion de « critique intersubjective transformatrice » [*transformative intersubjective criticism*, abréviation CIT]. Il s'agit du dialogue critique entre les membres d'une communauté scientifique, guidé par certaines normes⁶. Ce dialogue permet de préserver l'idéal

⁶ Cette idée est similaire à ce que Habermas a proposé comme éthique de la discussion, où des

d'objectivité, tout en reconnaissant la nature sociale des pratiques de production scientifique et la présence de valeurs sociales et culturelles en science (Longino, 1990, p. 62-63, 2002, p. 129).

Selon Longino, la CIT peut être obtenue en soumettant la communauté scientifique à quatre normes. Ces normes ont pour fonction de contraindre les structures et les interactions impliquées dans la production de connaissances en insistant sur l'importance du dialogue et de la critique (Longino, 1990, p. 76-80, 2002, p. 128-135; Ruphy, 2005 pour la traduction de certains termes clés trouvés dans les quatre normes) :

- 1) Il doit y avoir des forums où les méthodes, raisonnements, preuves observationnelles, etc., puissent être publiquement débattus et critiqués (*venues for criticism*), tels que le processus de révision par les pairs pour la publication. Les différents aspects de la recherche (la méthodologie, l'interprétation des données, etc.) doivent donc être examinés de manière critique par la communauté scientifique grâce à la disponibilité et à l'accessibilité de lieux de critiques.
- 2) La communauté doit présenter une réceptivité à la critique [*uptake of criticism*]. La critique doit être réellement entendue et doit pouvoir transformer la science au fil du temps (par contraste avec une simple tolérance à la critique). Par exemple, lorsqu'un article est soumis à publication et commenté par des évaluateurs externes, les commentaires doivent être intégrés, ce qui mènera à la modification du contenu de l'article, à une défense plus solide du propos ou à l'abandon de la thèse soutenue (voir Hicks, 2011 pour l'exemple).

participants délibèrent rationnellement à partir de leurs positions sociales réelles, en essayant de parvenir à un consensus, ou à l'idée de dialogue critique mise de l'avant par Mill ou Popper. La thèse de Longino est cependant plus spécifique, au sens où elle défend que ces normes représentent un idéal d'objectivité, l'objectivité sociale (voir Longino, 1990, chapitre 9, pour une discussion de la relation entre la position d'Habermas et la sienne).

- 3) La communauté doit posséder des standards publiquement reconnus [*shared public standards*] à partir desquels les théories, les hypothèses et les données sont évaluées. Les recherches sont critiquées (point 1) en fonction de différents critères épistémiques acceptés par une communauté. Les valeurs épistémiques qui guident la production de connaissances scientifiques (simplicité, pouvoir explicatif, etc.) font également partie de ces standards. Selon Longino, différentes sous-communautés épistémiques vont se dessiner en fonction de leur adhérence à certains standards et à certaines valeurs épistémiques. L'adéquation empirique est la seule valeur qui doit être communément partagée par tous les sous-groupes.
- 4) On doit retrouver une égalité d'autorité intellectuelle « tempérée » [*tempered equality of intellectual authority*] au sein de la communauté scientifique. Chacun des membres de la communauté doit pouvoir participer à la critique sans que son appartenance à un groupe socio-économique ou politique particulier décrédibilise l'expression de sa critique. Ainsi, la race, le genre, l'âge ou l'orientation sexuelle ne doivent pas venir interférer dans la participation aux processus critiques décrits plus haut. L'égalité de l'autorité intellectuelle est dite « tempérée » au sens où l'autorité intellectuelle doit être répartie en fonction de l'expertise uniquement. Malgré le critère de l'expertise, Longino précise que tout être humain doit être traité comme potentiellement capable de raisonner et de produire une critique (Longino, 2002, p. 131-133)⁷.

⁷ Les normes de l'OS ont de quoi rappeler l'éthique de la discussion développée par Habermas (1971), où un des participants provenant de diverses positions sociales participent à un processus délibératif rationnel pour tenter de parvenir à un consensus sur un sujet controversé. L'originalité de l'OS est d'importer ce type de mécanisme en sciences pour parvenir à une meilleure science. Pour Longino, les normes de l'OS permettent plus précisément d'identifier et de critiquer les biais individuels par la rencontre de points de vue divergents. Qui plus est, pour Longino, l'essentiel de ce processus délibératif n'est pas de parvenir à un consensus, mais plutôt de minimiser l'impact des biais et des valeurs problématiques sur la science.

Les hypothèses, théories et programmes de recherche peuvent être dits plus ou moins objectifs dans la mesure où la manière dont ils sont générés satisfait ces normes (Longino, 1990, p. 76). L'objectivité scientifique se présente donc par degrés, en ce que les communautés scientifiques mettent en place des procédures qui se conforment plus ou moins à ces quatre normes.

Dans le cadre de l'OS, la critique joue un rôle fondamental, puisque c'est par la différence de perspectives et de points de vue écoutés et entendus que la science peut progresser, en éliminant des présupposés d'arrière-plan qui se révéleraient problématiques. Pour Longino, une diversité de perspectives est nécessaire pour que soient mises en évidence les différentes présuppositions d'arrière-plan. Ceux et celles qui ne partagent pas les valeurs communément admises par une communauté épistémique sont plus à même d'identifier et de critiquer ces présupposés. Une communauté épistémique homogène court donc le risque de maintenir invisibles des présupposés d'arrière-plan problématiques (Longino, 1990, p. 28). Longino précise par ailleurs que plus une communauté scientifique est diversifiée, plus la CIT sera effective et générera des effets positifs pour la production de connaissances scientifiques. En effet, puisque les quatre normes ont pour but de rendre explicites les différents présupposés d'arrière-plan qui sous-tendent les hypothèses et théories scientifiques, ces présupposés peuvent être critiqués plus facilement. Par exemple, si des femmes et des personnes racisées sont intégrées dans les communautés scientifiques, les biais sexistes et racistes ont plus de chances d'être découverts et critiqués (voir Intemann, 2017 pour une discussion). Cette diversité peut aussi mener à davantage de créativité en termes de projets de recherche étant donné que ces projets sont généralement motivés par les intérêts et valeurs des chercheur.e.s (p. ex. : Solomon, 2006).

La conception de l'objectivité de Longino est donc en opposition avec la conception traditionnelle de la science et, par là, avec la conception traditionnelle de l'objectivité comme neutralité des valeurs. L'idéal de neutralité est ici remplacé par l'idéal de la

« gestion des valeurs sociales » [*social value management ideal*], (voir Rolin, 2017 à qui l'on doit cette appellation), où les valeurs non épistémiques ne doivent pas être éliminées, mais plutôt identifiées, discutées et critiquées. La théorie de l'OS développée par Longino rend possible une science influencée par les valeurs non épistémiques, tout en étant objective – une objectivité procédurale, par nature sociale.

4.7 Implications des critiques de l'idéal de la science indépendante des valeurs et de l'OS pour la définition du trouble mental

Comme démontré dans les chapitres précédents avec l'approche constructiviste et les critiques de la neutralité de la notion de dysfonction chez Boorse et Wakefield, il y a de bonnes raisons de penser que le concept de trouble mental est influencé par les valeurs. Dans le chapitre 2, j'ai aussi montré en quoi la psychiatrie pouvait être comprise comme une science influencée par les valeurs. Il apparaît clair désormais que les valeurs sociales et culturelles sont présentes en psychiatrie et dans la définition du concept de trouble mental. À la suite des précédentes sections de ce chapitre, il devrait également apparaître que l'idéal de la science indépendante des valeurs est à rejeter. Cependant, j'aimerais revenir à une intuition propre à l'approche NFO, selon laquelle l'influence des valeurs sociales et culturelles en science mènerait à une « mauvaise » science. Le fait de reconnaître la présence légitime de valeurs en science n'implique pas que cette présence soit *toujours* légitime. L'histoire litigieuse de la psychiatrie présentée en introduction et dans le chapitre 2 démontre en effet que certaines valeurs ont été et sont encore problématiques. Dans cette section, j'argumenterai donc qu'il faut se montrer suspect à l'égard de l'intrusion de certaines valeurs en psychiatrie, en utilisant deux contextes où la présence de valeurs a été (ou est toujours) préjudiciable :

la prévalence de diagnostic de trouble mental dans les populations racisées en raison de préjugés racistes et l'impact de stéréotypes sexistes sur le taux et le type de diagnostics de trouble mental chez les femmes. Dans ce contexte, j'entends par « racisme » et « sexisme » les préjugés dirigés contre certains groupes ou individus faisant partie de ces groupes et pour lesquels la différence identifiée chez ces groupes est marquée par l'infériorité (voir par exemple Nissim-Sabat, 2013, p. 139)⁸.

4.7.1 Racisme et trouble mental

De nombreux auteurs se sont intéressés aux liens entre le racisme et la psychiatrie (voir Metzl, 2009; Nissim-Sabat, 2009, 2013; Thomas et Sillen, 1972). L'exemple du diagnostic de drapétomanie, discuté dans l'introduction, constitue une illustration frappante du type de biais raciste qui peut intervenir en psychiatrie (voir Nissim-Sabat, 2009, p. 113). Rappelons que la drapétomanie, soit le fait pour les esclaves de tenter de fuir leur maître et de retrouver leur liberté, était considérée comme un trouble mental dans la deuxième moitié du 19^e siècle par le psychiatre Samuel Cartwright (1851/2004). Il est possible d'interpréter la pathologisation de ce comportement comme une manifestation de croyances racistes à l'égard des esclaves, qui étaient perçus comme déterminés à la servitude (« qualified for the trying duties of servitude », dans les mots

⁸ Il serait possible d'argumenter que plusieurs autres types de biais sont présents en psychiatrie. Prenons en exemple les doléances exprimées par le mouvement récent de la neurodiversité (voir Chapman, 2019). Ce mouvement soutient que les personnes neurotypiques sont perçues comme standard de normalité, tandis que les personnes neuroatypiques sont perçues comme anormales ou déviantes par rapport à ce standard. La crise de l'homosexualité discutée dans l'introduction peut également illustrer la présence de biais, homophobes cette fois, en psychiatrie, où l'esprit « normal » est associé à l'hétérosexualité.

de Cartwright, 2004, p. 32). Tenter de retrouver sa liberté était donc conçu comme une déviance vis-à-vis la nature servile et passive des esclaves et entraînait alors des symptômes de folie. Cartwright parle à cet effet de « barbarisme », de « superstition » et d'« ignorance », qui seraient associés aux esclaves affranchis (Cartwright, 2004, p. 38). Dans la vision raciste de Cartwright, il serait donc dans l'intérêt des esclaves de les priver de liberté, puisque leur nature, intrinsèquement différente de celle des blancs, ne pourrait pas le supporter. L'esclavage serait donc le remède à la folie des personnes noires placées en état de liberté (Cartwright, 2004, p. 39). Cet exemple est choquant et est historiquement situé dans le contexte du sud des États-Unis vers la deuxième moitié du 19^e siècle. Cependant, il présente l'avantage de démontrer de manière évidente comment des croyances racistes quant à une différence essentielle entre les personnes de couleurs noire et blanche peuvent entrer dans notre conception du trouble mental. L'intrusion de valeurs racistes peut alors présenter des effets extrêmement néfastes sur certaines populations, ici les personnes noires.

Un autre exemple plus récent des effets du racisme sur la classification psychiatrique peut être trouvé dans le travail de Rachel Cooper (2007). Cooper utilise des articles publiés dans *The American Journal for Psychiatry* entre 1844 et 1962 et distingue cinq contextes où des présuppositions racistes ont pu influencer les recherches psychiatriques :

- 1) Le choix du domaine/des axes de recherches;
- 2) La formulation des hypothèses à tester;
- 3) L'évaluation des éléments de preuve;
- 4) La présentation et l'analyse des résultats;
- 5) L'utilisation des résultats. (Cooper, 2007, p. 130-136)

Pour chacun de ces contextes, Cooper argue que des pré-supposés racistes ont mené à des effets épistémologiques pernicieux sur la recherche psychiatrique. Je ne vais pas

passer en revue l'ensemble de l'analyse qu'offre Cooper, mais uniquement les aspects qui concernent la conception du trouble mental. Dans un premier temps, Cooper affirme que « l'esprit blanc » a été implicitement considéré comme « l'esprit normal ». Cela se manifeste particulièrement lorsque la prévalence de diagnostic de trouble mental était différente entre les populations racisées et les populations blanches. Lorsque ces résultats étaient interprétés (4), cette différence était conçue comme une déviation par rapport au groupe « normal », soit celui de la population blanche (Cooper 2007, p. 135). Une autre illustration est la surreprésentation des populations noires par rapport à certains diagnostics. Cooper utilise une étude menée par Loring and Powell (1988, cités dans Cooper, 2007, p. 132), qui démontre que dans le cas du diagnostic de schizophrénie, les psychiatres étaient davantage prompts à poser le diagnostic auprès des patients noirs que des patients blancs (à ce sujet, voir aussi Metzl, 2009).

4.7.2 Sexisme et trouble mental

L'approche féministe a également permis de déceler la présence de présupposés sexistes sur le processus de catégorisation en psychiatrie, la recherche et les interventions cliniques (voir par exemple Boysen *et al.*, 2014; Boysen et Logan, 2017; Lunbeck, 1994; Martin, 2001; Potter, 2001; Shaw et Proctor, 2005; Showalter, 1985; Ussher, 1991). En lien avec la conception du trouble mental, Ussher a montré que des biais sexistes reliés à « l'esprit normal » avaient mené à un taux de diagnostic de certains troubles mentaux plus élevé chez les femmes, comme le trouble de la personnalité limite (Ussher, 1991, p. 71; il est à noter que des exceptions relatives à certaines catégories diagnostiques comme la schizophrénie existent et que l'intersection des identités « être noir.e » et « être une femme » peut augmenter la

possibilité de recevoir un diagnostic). Ussher explique cette surreprésentation par le fait que dans un contexte patriarcal, l'esprit normal ou sain a été associé à la masculinité. Ainsi, les femmes ont plus de chances d'être étiquetées comme « folles » que les hommes. Pour parvenir à cette conclusion, Ussher (1991, p. 160-170; 2011, p. 74-81) retrace le processus de catégorisation qui aurait participé à ce résultat et identifie deux possibles facteurs comme responsables : 1) la pathologisation des comportements exagérément féminins et 2) la pathologisation de la déviance par rapport aux stéréotypes féminins traditionnels.

En ce qui concerne le premier facteur, Ussher utilise le diagnostic passé de l'hystérie et le diagnostic actuel de trouble de la personnalité limite pour arguer que les symptômes associés à ces deux diagnostics, principalement la labilité ou la dérégulation émotionnelle, sont typiquement associés à l'infériorité des femmes. En ce qui concerne le deuxième facteur, la pathologisation de la déviance par rapport aux stéréotypes genrés, Ussher affirme que cela apparaît manifeste au regard de l'expression de la frustration ou de la rage chez les femmes (comme dans le cas du syndrome dysphorique prémenstruel, pour ce type d'analyse, voir aussi Caplan, 1995; Cosgrove et Riddle, 2004; Offman et Kleinplatz, 2004; Ussher *et al.*, 2000; d'un point de vue plus littéraire, des exemples peuvent être trouvés dans le roman *Howards End* d'E.M. Forster, 1910/1997), et du refus des rôles associés à la maternité (comme dans le cas de la dépression postnatale). Dans une lignée similaire, Potter examine le cas du trouble de la personnalité limite et la manière dont nos attentes genrées par rapport à la colère influencent le diagnostic (Potter, 2006, 2008). Le trouble de la personnalité limite se caractérise par plusieurs symptômes, dont celui de l'hostilité, soit le fait de présenter des épisodes de colères intenses et exagérées au regard des circonstances.

Selon Potter, nous sommes habitués à observer l'expression de la colère chez les hommes, mais moins chez les femmes. Cette habitude crée une normalisation de la colère chez les hommes, mais pas chez les femmes. Cette normalisation de la colère

masculine a pour effet de pathologiser la colère féminine. Par ailleurs, Potter ajoute que la perception de la colère pourra varier en fonction des genres. L'expression de la colère chez les hommes pourrait être interprétée comme assertive, alors qu'elle pourrait être interprétée comme agressive chez les femmes, et ainsi perçue comme pathologique. En ce sens, les femmes auraient tendance à recevoir davantage le diagnostic de trouble de la personnalité limite étant donné les attentes genrées que nous avons par rapport à la colère.

Ces exemples tendent donc à montrer, suivant l'analyse générale d'Ussher, qu'être « trop féminine », c'est-à-dire exprimer de manière extrême des traits liés à la féminité, ou trop se distancer de ces traits a mené à une augmentation de la fréquence selon laquelle les femmes reçoivent un diagnostic de trouble mental. Dans ce type d'analyse féministe, étant donné que la normalité est davantage associée aux comportements masculins ou aux respects des normes de genre, on peut affirmer que des biais sexistes ont pu affecter les pratiques diagnostiques envers les femmes et mener à un surdiagnostic des certains troubles mentaux chez les femmes.

Ces deux lignes d'analyses apportées par Cooper et Ussher et l'analyse des propos de Samuel Cartwright démontrent qu'il est en effet possible que la présence de valeurs au sein de la psychiatrie ait été et soit toujours problématique. En effet, les présupposés sexistes et racistes reliés au concept de trouble mental peuvent avoir des conséquences préjudiciables sur certains groupes traditionnellement socialement subordonnés. La démonstration des possibles effets nocifs des valeurs en science nous force à conserver l'intuition de la NFO selon laquelle la présence de normes sociales et culturelles en science peut être un problème. Cependant, puisque je considère les critiques de l'idéal de la science indépendante des valeurs convaincantes, il faut pouvoir maintenir la thèse de la légitimité des valeurs en science tout en proposant des moyens de contrôler cette influence. L'implication de mon analyse est plutôt que si ce type de valeurs n'est pas examiné, il peut conduire à des effets désastreux. La solution doit résider dans la

prévention des effets préjudiciables possiblement entraînés par les valeurs, et non pas dans leur élimination totale (dans plusieurs domaines scientifiques, cette tâche n'est ni réaliste ni souhaitable, comme le montrent les critiques de l'idéal de la science indépendante des valeurs). Le modèle de l'objectivité sociale développé par Longino me semble pouvoir répondre à cet objectif. Je montrerai comment dans le chapitre 6.

Conclusion

Dans ce chapitre, je me suis intéressée à une présupposition sous-jacente à l'approche NFO : l'idéal de la science indépendante des valeurs. J'ai défendu qu'il y avait de bonnes raisons de s'opposer à la poursuite de cet idéal et j'ai présenté une conception alternative de l'objectivité scientifique : l'objectivité sociale. J'ai ensuite montré comment l'objectivité sociale peut trouver sa place en psychiatrie en analysant la manière dont certaines valeurs répudiables quant au trouble mental pouvaient affecter négativement les individus appartenant à des groupes socialement subordonnés.

Références du chapitre 4

- Belkhir, J. (1994). Race, sex, class & “intelligence” scientific racism, sexism & classism. *Race, Sex & Class*, 1(2), 53-83.
- Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, 3(2), 207-220.
- Biddle, J. (2013). State of the field : Transient underdetermination and values in science. *Studies in History and Philosophy of Science Part A*, 44(1), 124-133.
- Biddle, J. et Kukla, R. (2017). The geography of epistemic risk. Dans K. C. Elliott et T. Richards (dir.), *Exploring inductive risk : Case studies of values in science* (p. 215-238). New York : Oxford University Press.
- Boorse, C. (1997). A rebuttal on health. Dans J. M. Humber et R. F. Almeder (dir.), *What is disease?* (p. 1-134). Totowa, NJ : Humana Press.
- Boysen, G. A. et Ebersole, A. (2014). Expansion of the concept of mental disorder in the DSM-5. *The Journal of Mind and Behavior*, 35(4), 225-243.
- Boysen, G., Ebersole, A., Casner, R. et Coston, N. (2014). Gendered mental disorders : Masculine and feminine stereotypes about mental disorders and their relation to stigma. *The Journal of Social Psychology*, 154(6), 546-565.
- Brown, Matthew J. (2013). Values in science beyond underdetermination and inductive risk. *Philosophy of Science*, 80(5), 829-839.
- Caplan, P. J. (1995). *They say you're crazy : How the world's most powerful psychiatrists decide who's normal*. Caplan, MA : Addison-Wesley.
- Cartwright, S. (1851/2004). Report on the diseases and physical peculiarities of the Negro race. Dans A. L. Caplan, J. J. McCartney et D. A. Sisti (dir.), *Health*,

disease, and illness : Concepts in medicine (p. 28-39). Washington : Georgetown University Press.

Cooper, R. (2007). *Psychiatry and philosophy of science*. Londres : Routledge.

Cosgrove, L. et Riddle, B. (2004). Gender bias and sex distribution of mental disorders in the DSM-IV-TR. Dans P. J. Caplan et L. Cosgrove (dir.), *Bias in psychiatric diagnosis* (p. 127-140). Lanham, MA : Jason Aronson :.

Cranor, C.F. (1990). Some moral issues in risk assessment. *Ethics*, 101(1), 123-143.

Crasnow, S. (2001). Models and reality : When science tackles sex. *Hypatia*, 16(3), 138-148.

Crasnow, S. (2013). Feminist philosophy of science : Values and objectivity. *Philosophy Compass*, 8(4), 413-423.

Daston, D. et Galison, P. (2007). *Objectivity*. Cambridge, MA : MIT Press.

Dorato, M. (2004). Epistemic and non-epistemic values in science. Dans P. Machamer et G. Wolters (dir.), *Science, values, and objectivity* (p. 52-77). Pittsburgh : University of Pittsburgh Press.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, 67(4), 559-579.

Douglas, H. (2003). The moral responsibilities of scientists (tensions between autonomy and responsibility). *American Philosophical Quarterly*, 40(1), 59-68.

Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, 138(3), 453-473.

Douglas, H. (2007). Rejecting the ideal of value-free science. Dans H. Kincaid, J. Dupré et A. Wylie (dir.), *Value-free science? Ideals and illusions* (p. 120-141). New York : Oxford University Press.

- Douglas, H. (2008). The role of values in expert reasoning. *Public Affairs Quarterly*, 22(1), 1-18.
- Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh : University of Pittsburgh Press.
- Dupré, J. (2007). Fact and value. Dans H. Kincaid, J. Dupré et A. Wylie (dir.), *Value-free science? Ideals and illusions* (p. 27-41). New York : Oxford University Press.
- Elliott, K. C. (2011). *Is a little pollution good for you? Incorporating societal values in environmental research*. New York : Oxford University Press.
- Feyerabend, P. (1962). Explanation, reduction and empiricism. Dans H. Feigl et G. Maxwell (dir.), *Scientific explanation, space, and time* (p. 28-97). Mineapolis : University of Minneapolis Press.
- Forster, E. M. (1910/1997). *Howards end*. New York : Macmillan.
- Gross, P. R. et Levitt, N. (1997). *Higher superstition : The academic left and its quarrels with science*. Baltimore : Johns Hopkins University Press.
- Haack, S. (1993). Epistemological reflections of an old feminist. *Reason Papers*, 18, 31-43.
- Haack, S. (2003). Knowledge and propaganda. Dans C. Pinnick, N. Koertge et R. Almeder (dir.), *Scrutinizing feminist epistemology : An examination of gender in science* (p. 7-19). Piscataway, NJ : Rutgers University Press.
- Haely, K.C. (2008). *Objectivity in the feminist philosophy of science*. Londres : A&C Black.
- Haraway, D. (1988). Situated knowledges : The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575-599.
- Harding, S. G. (1986). *The science question in feminism*. Ithaca, NY : Cornell University Press.

- Hicks, D. (2011). Is Longino's conception of objectivity feminist? *Hypatia*, 26(2), 333-351.
- Howalter, E. (1987). *The female malady*. Londres : Virago.
- Intemann, K. (2005). Feminism, underdetermination, and values in science. *Philosophy of Science*, 72(5), 1001-1012.
- Intemann, K. (2017). Feminism, values, and the bias paradox. Dans K. C. Elliott et D. Steel (dir.), *Current controversies in values and science* (p. 130-131). New York : Routledge.
- Kendell, R. E. (1986). What are mental disorders? Dans M. Freedman, R. Brotman et I. Silverman (dir.), *Issues in psychiatric classification* (p. 23-45). New York : Human Sciences Press.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. (2^e édition). Chicago : University of Chicago Press.
- Lloyd, E. (1995). Objectivity and the double standard for feminist epistemologies. *Synthese*, 104(3), 351-381.
- Longino, H. (1983). Beyond "bad science" : Skeptical reflections on the value-freedom of scientific inquiry. *Science, Technology & Human Values*, 8(1), 7-17.
- Loring, M. et Powell, B. (1988). Gender, race, and DSM-III : A study of the objectivity of psychiatric diagnostic behavior. *Journal of Health and Social Behavior*, 29(1), 1-22.
- Lunbeck, E. (1994). *The psychiatric persuasion : Knowledge, gender, and power in modern America*. Princeton, NJ : Princeton University Press.
- Martin, E. (1991). The egg and the sperm : How science has constructed a romance based on stereotypical male-female roles. *Signs : Journal of Women in Culture and Society*, 16(3), 485-501.

- Martin, N. (2001). Feminist bioethics and psychiatry. *The Journal of Medicine and Philosophy*, 26(4), 431-441.
- Metzl, J. M. (2009). *The protest psychosis : How schizophrenia became a black disease*. Boston : Beacon Press.
- Nagel, T. (1986). *The view from nowhere*. New York : Oxford University Press.
- Nissim-Sabat, M. (2009). *Neither victim nor survivor : Thinking toward a new humanity*. Plymouth : Lexington Books.
- Nissim-Sabat, M. (2013). Race and gender in philosophy of psychiatry : Science, relativism, and phenomenology. Dans K. W. M. Fulford, M. Davies, G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini et T. Thornton (dir.), *The Oxford Handbook of Philosophy and Psychiatry* (p. 139-158). Oxford : Oxford University Press.
- Potter, E. (2006). *Feminism and philosophy of science : An introduction*. New York : Routledge.
- Potter, N. N. (2006). What is manipulative behavior, anyway? *Journal of Personality Disorders*, 20, 139-156.
- Potter, N. N. (2008). The problem with too much anger : A philosophical approach to understanding anger in borderline personality disordered patients. Dans C. Charland et P. Zachar (dir.), *Fact and value in emotion* (p. 53-86). Amsterdam : John Benjamins.
- Putnam, H. (2002). *The collapse of the fact/value dichotomy*. Cambridge, MA : Harvard University Press.
- Richards, J. R. (1995). Why feminist epistemology isn't (and the implications for feminist jurisprudence). *Legal Theory*, 1(4), 365-400.
- Roberts, D. (2013). Thick concepts. *Philosophy Compass*, 8(8), 677-688.

- Rolin, K. (2017). Can social diversity be best incorporated into science by adopting the social value management ideal? Dans K. C. Elliott et D. Steel (dir.), *Current controversies in values and science* (p. 113-129). New York : Routledge.
- Ruphy, S. (2015). Rôle des valeurs en science : Contributions de la philosophie féministe des sciences. *Écologie & politique*, 2, 46-47.
- Ruphy, S. (2006). “Empiricism all the way down” : A defence of the value-neutrality of science in response to Helen Longino’s contextual empiricism. *Perspectives on Science*, 14(2), 189-214.
- Shaw, C. et Proctor, G. (2005). Women at the margins : A critique of the diagnosis of borderline personality disorder. *Feminism & Psychology*, 15(4), 483-490.
- Shields, S. (1975). Functionalism, darwinism, and the psychology of women. *American Psychologist*, 30(7), 739-754.
- Shields, S. A. (1982). The variability hypothesis : The history of a biological model of sex differences in intelligence. *Signs : Journal of Women in Culture and Society*, 7(4), 769-797.
- Shields, S. A. (2007). Passionate men, emotional women : Psychology constructs gender difference in the late 19th century. *History of Psychology*, 10(2), 92.
- Solomon, M. (2006). Norms of epistemic diversity. *Episteme*, 3(1), 23-36.
- Stanford, K. (2017). Underdetermination of scientific theory. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/spr2017/entries/scientific-underdetermination/>
- Thomas, A. et Sillen, S. (1972). *Racism and psychiatry*. New York : Carol Publishing Group.
- Turnbull, M. G. (2018). Underdetermination in science : What it is and why we should care. *Philosophy Compass*, 13(2). doi.org/10.1111/phc3.12475

Tyrer, P. et Steinberg, D. (2006). *Models for mental disorder*. West Sussex : John Wiley & Sons.

Ussher, J. M. (1991). *Women's madness : Misogyny or mental illness?* Londres : Harvester Wheatsheaf.

Wakefield, J. C. (2007). The concept of mental disorder : Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, 6(3), 149-156.

Williams, B. (1985). *Ethics and the limits of philosophy*. Cambridge, MA : Harvard University Press.

CHAPITRE 5

LIMITES DE L'OBJECTIVITÉ SOCIALE ET PROPOSITIONS D'AMENDEMENTS

Introduction

Dans ce chapitre, je me pencherai sur les limites que peut rencontrer la théorie de l'objectivité sociale de Longino. Bien que ma thèse concerne la définition du trouble mental, dans ce chapitre, je m'éloignerai du contexte de la psychiatrie pour me concentrer uniquement sur la conception de l'objectivité sociale de Longino et sa réception critique en philosophie des sciences. Ce détour est motivé par le fait que si l'objectivité sociale de Longino ne résiste pas à ses critiques en philosophie des sciences, ma thèse s'en trouvera affaiblie puisqu'elle est basée principalement sur cette théorie. En fait, ce chapitre me permettra non seulement de montrer que l'OS n'est pas ébranlée par les tirs qu'elle a eu à affronter de la part de ses critiques, mais également d'approfondir et de préciser plusieurs aspects de l'OS. Cette discussion me donnera ainsi la possibilité de répondre à des critiques qui ne sont pas légitimes, mais aussi d'amender cette théorie dans les cas où les réponses possibles sembleraient insatisfaisantes. À la fin de ce chapitre, j'aboutirai à une conception de l'OS plus solide, ce qui me mènera à appliquer des principes éprouvés de l'objectivité sociale au cas de la définition du trouble mental dans les deux chapitres suivants.

Dans les écrits, j'ai identifié cinq types de critiques qui pourraient poser problème dans le cadre de ma thèse. Dans la section 5.1, j'aborderai la critique, pour le moins surprenante, d'antiféminisme qui a été adressée à Longino. Dans la section 5.2, je me pencherai sur le type de diversité requis au sein d'une communauté épistémique dite « socialement objective ». Je mets en relation la théorie féministe du *standpoint* (pour ce terme, je conserve la formulation anglaise qui a un sens technique dans les écrits, mais on pourrait le comprendre comme « une situation sociale ») avec celle de l'objectivité sociale de Longino et je tente de voir si ces deux théories peuvent converger. Dans la section 5.3, j'interrogerai la notion de « communauté épistémique » au sein de la théorie de l'objectivité sociale. J'y étudie une proposition intéressante selon laquelle des membres faisant partie de communauté non scientifique pourraient être intégrés à une communauté épistémique « socialement objective ». Dans la section 5.4, je m'intéresserai à la critique de relativisme qui a été adressée à Longino. Pour finir, dans la section 5.5, je poserai la question de la justesse de l'analyse conceptuelle de Longino au regard du concept général d'objectivité scientifique. Comme nous le verrons, ces différentes critiques ne remettent pas en cause l'OS, mais demandent cependant à ce que certains amendements soient faits pour que la théorie demeure convaincante.

5.1 L'OS est-elle antiféministe?

Comme présenté dans le chapitre 4, la théorie de l'objectivité sociale de Longino prend ancrage, du moins en partie, dans le courant de la philosophie féministe des sciences. Or récemment, certaines chercheuses ont accusé Longino de n'avoir pas su offrir une

théorie de l'objectivité pleinement féministe; elles ont même affirmé que celle-ci aurait des implications allant contre les principes féministes. Je m'intéresse à cette critique parce qu'elle permet de préciser le type de perspectives requises pour qu'une communauté scientifique soit dite objective selon les standards de Longino, précision qui sera utile pour son application à la définition du trouble mental. L'accusation d'antiféminisme ou de non-féminisme a pris plusieurs formes. De manière générale, il s'agit du souci selon lequel les valeurs et intérêts féministes ne pourraient pas acquérir de statut particulier puisque la théorie de l'objectivité sociale requiert la présence d'une diversité de perspectives et d'intérêts. En ce sens, Longino ne fournirait pas d'arguments en faveur de bénéfices particuliers apportés par le point de vue féministe par rapport à d'autres (Intemann, 2008, 2011). Une autre formulation de cette critique cible la relativité morale de la théorie de Longino : puisqu'aucune valeur morale n'est supérieure à une autre, les valeurs féministes ne seraient pas supérieures aux valeurs sexistes. La théorie de Longino impliquerait donc une forme de relativisme moral. Hicks (2011) a poussé l'accusation plus loin en arguant que par le critère d'égalité d'autorité intellectuelle tempérée, l'une des quatre normes de l'OS, la théorie de Longino demanderait la recherche active de l'expression de plusieurs perspectives au sein de la recherche scientifique, y compris des perspectives racistes et sexistes. Si Hicks a raison, l'OS semble problématique dans un projet comme le mien. Comme expliqué dans le chapitre 4, cette norme implique que les perspectives des femmes et des autres groupes marginalisés soient recherchées pour favoriser l'explicitation et la critique des présupposés d'arrière-plan contenus dans les théories scientifiques. Cependant, selon l'interprétation de Hicks, pour maximiser la diversité, les perspectives des personnes racistes et misogynes devraient être elles aussi recherchées, puisque c'est par la multiplication des points de vue possibles que les présupposés d'arrière-plan problématiques pourront être identifiés. Ainsi, même si l'objectif de Longino est de maximiser la diversité épistémique en redonnant voix aux personnes faisant partie de groupes opprimés traditionnellement exclus de la production du savoir

scientifique, sa conception de l'objectivité demanderait également de favoriser l'apport critique de groupes oppresseurs tels que les personnes racistes et sexistes. Autrement dit, la conception de l'objectivité de Longino demanderait à intégrer activement et explicitement des groupes oppresseurs au sein de la recherche scientifique puisque ces perspectives pourraient elles aussi servir à mettre au jour des présupposés problématiques. L'OS de Longino aurait donc pour implication la promotion de valeurs et d'intérêts à la fois féministes *et* antiféministes. Puisque les valeurs féministes demandent minimalement l'identification et la lutte contre l'oppression des groupes marginalisés (Borgerson, 2011), et que l'OS semble demander la présence de valeurs antiféministes, il n'est plus évident que l'OS puisse se prétendre féministe.

En réponse à cette critique, Rolin (2017) a proposé la solution suivante : des attitudes sexistes ou racistes entrent directement en conflit avec deux des normes de l'OS, soit le critère de réceptivité à la critique et le critère d'égalité d'autorité intellectuelle tempérée, si celui-ci est compris correctement. Parce que les idéologies racistes ou sexistes impliquent que ce ne sont pas tous les êtres humains qui sont perçus comme capables de participer également à la critique en fonction de leur expertise, elles sont en contradiction avec le critère d'égalité d'autorité intellectuelle tempérée. Rappelons que pour Longino, ce critère implique deux éléments. D'abord, cette égalité d'autorité intellectuelle est partagée en fonction de l'expertise de chacun. Les scientifiques possédant une expertise dans un domaine de recherche spécifique sont considérés comme détenant plus de connaissances dans ce domaine. Cependant, pour Longino, l'autorité intellectuelle n'est pas seulement une question de connaissances supérieures. Il s'agit également de posséder des capacités cognitives telles que des habiletés observationnelles, d'analyse et de réflexion, qui rendent possible l'expression de commentaires pertinents et convaincants sur la recherche menée. Des croyances racistes et sexistes empêcheraient d'atteindre ce critère d'objectivité, étant donné qu'elles rejettent la participation d'un individu à la production de savoir scientifique

sur la base de préjugés sociaux, et non sur la base de l'autorité intellectuelle. Qui plus est, parce que les idéologies racistes ou sexistes décrédibilisent directement ou indirectement certains individus en fonction de leur appartenance à des groupes sociaux marginalisés, elles n'assurent pas une réceptivité suffisante à la critique de la part de la communauté scientifique. Cet élément constitue une autre des quatre normes de l'OS proposées par Longino. Ainsi, suivant la réponse de Rolin (2017), Longino posséderait les ressources pour parer cette attaque si les normes de réceptivité à la critique et d'égalité d'autorité intellectuelle tempérée sont interprétées correctement.

Cependant, un élément critique semble pouvoir persister quant à la possibilité d'idéologies antiféministes au sein d'une science « socialement objective » : bien que les critères proposés par Longino soient *formellement* incompatibles avec des attitudes racistes ou sexistes, dans la pratique, il est possible que ces critères ne permettent pas réellement d'exclure ces idéologies de la communauté scientifique. À elles seules, les normes de l'OS pourraient ne pas être suffisantes pour mettre en place des mécanismes d'exclusion effectifs des attitudes racistes ou sexistes. Pour répondre à cette préoccupation plausible, il faut vraisemblablement admettre qu'il est en effet nécessaire d'élaborer des mécanismes effectifs qui préviennent l'intrusion d'attitudes racistes ou sexistes dans les communautés scientifiques. Or, l'ajout de ces mécanismes ne semble pas venir contredire la conception de l'OS défendue par Longino, mais demande plutôt qu'on en précise les mécanismes de fonctionnement. Ces précisions viendraient spécifier comment, dans la pratique, les attitudes racistes ou sexistes doivent être contrées. Le cadre de cette thèse ne me permet pas de discuter en détail les différents moyens concrets qui pourraient être employés pour lutter contre les attitudes racistes ou sexistes. Notons tout de même que différentes solutions existent dans les écrits (p. ex. : anonymisation, discrimination positive, etc.) et que ces outils employés pour lutter contre les attitudes racistes ou sexistes ne semblent pas être en conflit avec les normes de l'OS, si elles sont interprétées comme le suggère Rolin.

5.2 Quelle diversité pour la communauté épistémique? La théorie de la connaissance située contre l'empirisme féministe?

Dans un article de 2011, la philosophe de la médecine Kristen Borgerson a soutenu que, dans le cadre de la théorie de l'OS de Longino, la manière dont une communauté épistémique devait être diversifiée demandait à être précisée. Cette précision est importante pour ma thèse, puisqu'il est essentiel de bien comprendre la notion de diversité pour pouvoir l'appliquer au cas de la définition du trouble mental. La critique de Borgerson est en phase avec un article récent de Steel et ses collègues (2018) qui présentent une catégorisation des différents concepts de diversité présents dans les écrits en philosophie des sciences. Ces conceptions de la diversité auraient différentes implications sur les plans épistémique et éthique (Steel *et al.*, 2018). Selon ces auteurs, la diversité peut faire référence à des positions sociales, des valeurs et des intérêts, des orientations théoriques, des habiletés, etc. (voir aussi Rolin, 2011 pour le même type d'analyse). Selon Borgerson, pour la théorie de l'objectivité sociale de Longino, la diversité de la communauté épistémique devrait être celle qui va permettre de mettre en lumière les présuppositions d'arrière-plan contenues dans les théories et hypothèses scientifiques. Selon son interprétation, la diversité participe à l'OS dans la mesure où elle permet d'intégrer des perspectives possiblement contradictoires. Des perspectives en désaccord permettront alors d'identifier et de critiquer de manière plus efficace les présuppositions d'arrière-plan qu'une diversité « aveugle », qui ne prendrait pas en compte l'importance du désaccord. Borgerson propose d'ailleurs de renommer la quatrième norme de l'objectivité sociale, l'égalité d'autorité intellectuelle tempérée, comme « l'encouragement des perspectives divergentes » (*Cultivation of Diverse*

Perspectives). Cette reformulation aurait pour conséquence d'attirer l'attention sur le type de diversité attendu pour qu'une communauté épistémique soit dite « socialement objective », en mettant l'accent sur l'importance du désaccord. Or ce type de diversité ne fait pas consensus.

Dans un article de 2010, la philosophe féministe des sciences Kristen Intemann a mené une analyse des différences et des similitudes existantes entre l'approche de l'empirisme féministe (à laquelle Longino appartient; cette approche est également défendue aussi par Miriam Solomon [2001, 2006], Lynn Hankinson Nelson [1990] et Sharoy Clough [2003]) et la théorie féministe du *standpoint* (comme défendue par Sandra Harding [1986, 2015] et Alyson Wylie [2003], entre autres). Alors que Sandra Harding avait distingué l'empirisme féministe de la théorie féministe du *standpoint* dans les années 1980 (Harding, 1986), Intemann soutient que ces deux théories devraient converger pour former une théorie hybride qu'elle nomme la théorie féministe et empiriste du *standpoint* (*feminist standpoint empiricism*). Cette convergence aurait l'avantage de fournir des bases théoriques permettant de légitimer une diversité de *standpoints*, et non pas une diversité de point de vue et d'intérêts, comme le soutiennent Longino et Borgerson. Pour mieux comprendre la proposition d'Intemann, je vais d'abord exposer brièvement la théorie féministe du *standpoint* et la manière dont la diversité y est comprise. Je vais ensuite discuter la thèse selon laquelle la diversité d'une communauté épistémique devrait se manifester en termes de *standpoints* dans une conception de l'objectivité sociale. Je me permets une plus longue présentation de la théorie féministe du *standpoint* étant donné que cette théorie me semble pertinente pour penser le cas de la définition du trouble mental.

5.2.1 Présentation de la théorie féministe du *standpoint*

La théorie féministe du *standpoint* prend ancrage dans les travaux d'épistémologie marxiste portant sur le point de vue (*standpoint*) du prolétariat par rapport à la bourgeoisie. Selon ces travaux, des inégalités entre les classes créent des points de vue différents sur le monde. Puisque ce sont les classes supérieures bourgeoises qui instituent des cadres conceptuels et méthodologiques particuliers pour étudier le monde, ces cadres représentent les intérêts de ces classes, aux dépens des classes inférieures. Les classes inférieures, pour pouvoir fonctionner en société, doivent intégrer ces cadres conceptuels et méthodologiques. Or, les individus appartenant à ces classes inférieures ont aussi une expérience qui ne se conforme pas avec ces cadres en raison de leur situation d'oppression. En raison de cette double perspective (celle de la bourgeoisie et celle de leur groupe social opprimé), les classes inférieures sont alors plus à même de percevoir la partialité des cadres conceptuels et méthodologiques dominants puisqu'ils ne s'appliquent pas à leurs expériences. Interprétée dans une perspective féministe, cette théorie peut présenter d'autres implications : le point de vue des femmes peut également permettre de critiquer le point de vue dominant, ici masculin (Harding, 2015). Intemann (2010), s'inspirant des travaux d'Alison Wylie (2003), caractérise la théorie féministe du *standpoint* ainsi :

- 1) La thèse de la connaissance située : notre situation influence de manière systématique notre connaissance.
- 2) La thèse de l'avantage épistémique : certaines situations sociales, particulièrement celles des groupes opprimés ou marginalisés, présentent un avantage épistémique dans certains contextes.

La thèse de la connaissance située implique que la situation dans laquelle se trouve un individu a un impact sur sa connaissance. Anderson résume cette thèse ainsi :

These kinds of situatedness affect knowledge in several ways. They influence knowers' access to information and the terms in which they represent what they know. They bear on the form of their knowledge (articulate/implicit, formal/informal, by acquaintance or description, and so forth). They affect their attitudes toward their beliefs (certainty/doubt, dogmatic/open to revision), their standards of justification (relative weights they give to different epistemic values such as predictive power and consilience, amount, sources, and kinds of evidence they require before they accept a claim, etc.), and the authority with which they lay claim to their beliefs and can offer them to others. Finally, they affect knowers' assessment of which claims are significant or important (Anderson, 2019).

Nos différentes perspectives sur le monde influenceront donc le type de connaissances que nous avons et sa qualité/profondeur. Par exemple, quelqu'un qui fait l'expérience de la douleur possède une connaissance de la douleur à la première personne, alors qu'un physicien qui étudie la douleur sur le plan neurophysiologique a une connaissance de la douleur à la troisième personne. Un voleur peut avoir une connaissance d'un cadenas qui comprend les différentes méthodes qui permettent de mieux l'ouvrir, alors que la connaissance d'un propriétaire de vélo consiste plutôt à se le représenter comme un objet dont la fonction est de protéger son bien (la majorité des propriétaires de vélo ne savent pas exactement comment briser un cadenas). Dans ce cas-ci, la profondeur et la qualité de la connaissance varient sur un même objet. Un sans-abri pourrait avoir une connaissance d'une université selon ses propres intérêts, en fonction des ressources que cette université peut lui offrir (p. ex. : toilettes, lieu.x pour dormir et se réchauffer, horaire de patrouille des agents de sécurité, etc.). Par contraste, une étudiante pourrait avoir une connaissance différente de cette même université parce que ses intérêts diffèrent de ceux du sans-abri (p. ex. : le lieu de son bureau de travail, les horaires du centre sportif, etc.). Dans le contexte de la psychiatrie,

on peut penser à la connaissance qu'auraient un patient et un psychiatre d'un traitement : le patient aura généralement une connaissance expérientielle des effets (positifs et/ou négatifs) du traitement, ainsi qu'une connaissance des motivations personnelles et des besoins qui poussent à chercher un traitement, alors que les chercheurs en psychiatrie auront généralement une connaissance plus descriptive du traitement, souvent centrée sur les aspects neurobiologiques, génétiques et/ou épidémiologiques (voir Cuijpers, 2011 sur le trouble de dépression majeur). C'est donc en ce sens que l'on parle de « connaissance située » (Anderson, 2019) : notre situation affecte notre manière de connaître le monde, et cela aboutit à des connaissances différentes qui varient selon ces situations.

Dans le contexte de la théorie féministe du *standpoint*, l'idée de situation est analysée sous le prisme du social. Dans ce cadre, le concept de *standpoint* est utilisé au sens métaphorique. Les individus occuperaient des situations sociales, culturelles ou politiques différentes s'accompagnant de relations de pouvoir particulières (Crasnow, 2013). Le *standpoint* d'un individu est induit par l'identité sociale et les rôles sociaux que doit jouer celui-ci (le genre, le faire d'être membre d'un groupe racisé, l'orientation sexuelle, l'occupation, l'allégeance politique, etc.). Un *standpoint* peut être à l'intersection de plusieurs groupes sociaux menant à une pluralité de situations sociales, comme le fait d'être une femme noire lesbienne (Collins, 1986; Harding, 2015). Dans la théorie féministe du *standpoint*, cette situation sociale particulière a un impact sur notre manière de connaître. Plus précisément, on y affirme que le point de vue socialement situé des femmes a un impact sur ce qu'elles connaissent du monde, principalement en raison des structures de domination patriarcale et des idéologies sexistes. Par exemple, dans un contexte domestique traditionnel, une femme pourra percevoir la poussière ou le désordre dans sa maison alors que son mari ne remarquera rien. Cette différence de perspectives peut s'expliquer par une socialisation différenciée selon le genre et une division genrée du travail (la socialisation genrée traditionnelle

implique que les femmes s'occupent des tâches domestiques et que les hommes occupent un emploi à l'extérieur de la maison). Un autre exemple pourrait être celui d'une interaction hétérosexuelle où un homme fait des avances à une femme et la femme répond par un sourire. Dans ces circonstances, l'homme pourra interpréter le sourire comme un signe d'intérêt, alors qu'une femme qui observerait la scène de l'extérieur pourrait percevoir le sourire comme une tentative gênée d'échapper aux avances de l'homme. Ici, la perception de la situation est informée par des expériences de séduction qui diffèrent pour chacun des genres (Anderson, 2019). Ainsi, dans le cadre de la théorie féministe du *standpoint*, la thèse de la connaissance située se traduit par l'idée de *standpoint* : nos positions *sociales* ont un impact sur nos connaissances, principalement sur les connaissances influencées par les structures de pouvoir relatives à notre *standpoint*.

Déoulant de la thèse de la connaissance située (Crasnow, 2018), la thèse de l'avantage épistémique vient spécifier trois choses : le *standpoint* qui est à même de produire une meilleure connaissance sur le monde, la façon dont un tel privilège épistémique émerge et l'étendue de ce privilège (Anderson, 2019). Cette thèse met de l'avant l'idée selon laquelle les groupes socialement opprimés ou marginalisés, en raison de leur situation sociale défavorisée, pourraient avoir un avantage épistémique par rapport à des personnes qui sont en situation de privilège (p. ex. : les hommes, les personnes blanches, etc.). Ce privilège épistémique s'expliquerait par l'expérience des individus en situation sociale d'oppression : leur *standpoint* leur donnerait une connaissance par expérience de cette oppression, leur permettant de porter attention à des choses différentes par rapport au point de vue dominant et de mieux connaître certains aspects de l'oppression (Wylie, 2003). Cette thèse s'inspire des travaux de la sociologue Patricia Hill Collins (1986), qui défend que les chercheur.e.s provenant de groupes marginalisés ont une perspective d'initié/étranger (*insider/outsider*). La formation des initiés/étrangers leur donne un cadre théorique et conceptuel conforme à leur discipline

(perspective d'initié), mais leurs expériences en tant qu'étrangers suggèrent que ce cadre n'est peut-être pas approprié pour rendre compte des phénomènes à l'étude. Cette distance entre la formation d'initié et les expériences d'étranger confère à ce type d'individu un avantage épistémique : celui de pouvoir identifier les limites méthodologiques ou théoriques d'une discipline. Le fait de bénéficier d'une formation de scientifique et de provenir d'un groupe marginalisé permet d'identifier des présupposés qui apparaissent invisibles aux individus appartenant à des groupes privilégiés. L'avantage épistémique provient donc des expériences qui sont associées à une situation sociale marginalisée, expériences qui pourront influencer la connaissance en permettant de rendre visible et critiquable ce qui est communément admis par un groupe non marginalisé. Ces éléments peuvent être de l'ordre des preuves à considérer, des phénomènes pertinents à étudier, de la méthodologie utilisée, etc. (Wylie, 2012; voir aussi Crasnow, 2018). À cet effet, Wylie (2003) écrit :

And whenever commonalities of location and experience give rise to critical (oppositional) consciousness about the effects of social location, it may be possible to identify a distinctive standpoint to which strategic epistemic advantage accrues, particularly in grasping the partiality of a dominant way of thinking, bringing a new angle of vision to bear on old questions and raising new questions for empirical investigation.

Le recul que permet la position d'étranger (le *standpoint* marginalisé) par rapport au cadre dominant constitue donc un avantage épistémique.

Il est important de noter que, dans le cadre de la recherche, l'avantage épistémique dont il est question est celui de pouvoir mettre en lumière des présuppositions infondées par rapport au groupe marginalisé auquel l'individu appartient ainsi que d'exprimer un certain nombre de préoccupations dudit groupe qui sont ignorées par le groupe dominant. En ce sens, le fait d'être en situation sociale marginalisée ne donne pas d'avantage épistémique dans tous les domaines de connaissance et n'est pas

automatique. Comme l'illustre Wylie (2003) à l'aide d'un exemple, un employé d'usine du District Maquiladoras possède une connaissance accrue des politiques managériales pour produire le plus de profit au sein d'une entreprise, mais pas nécessairement des connaissances par rapport aux mouvements internationaux de capitaux ou à l'impact général des relocalisations d'usines en Thaïlande. Dans le cadre de la théorie féministe du *standpoint*, le genre est l'une des dimensions sociales qui peuvent conférer un avantage épistémique. Suivant le caractère non généralisable du privilège épistémique, celui-ci est circonscrit aux présuppositions et aux pratiques sexistes qui peuvent s'infiltrer au sein d'une discipline. Autrement dit, les femmes sont plus à même d'identifier des présuppositions sexistes en raison de leur position sociale de subordonnées (Anderson, 2019). Par exemple, les femmes peuvent avoir une meilleure connaissance des inégalités liées au travail domestique et à la maternité. Par ailleurs, il est également important de préciser que la théorie de la connaissance située ne se base pas sur des postulats essentialistes (Crasnow, 2006; Wylie, 2003) : les *standpoints* ne sont pas fixes et immuables, et les membres qui appartiennent à des groupes marginalisés n'auront probablement pas une perspective unique sur le monde. Qui plus est, il convient d'insister à nouveau sur le fait que le *standpoint* d'un individu peut être à l'intersection de plusieurs groupes subordonnés (p. ex. : classe, couleur de peau, neurotypicité/atypicité, orientation sexuelle, etc.) (Intemann, 2016). L'avantage épistémique est donc contingent et dépend des structures sociales en place à un moment donné. Par exemple, si le racisme cessait d'exister, le fait d'avoir une couleur de peau particulière n'aurait pas le même impact sur ce qui est connu actuellement par quelqu'un qui est racisé (Intemann, 2016). L'idée de *standpoint* n'implique donc pas une perspective unique et homogène sur le monde, et l'impact épistémique de cette situation sociale est amené à changer en fonction des mutations sociales que connaîtra une société.

En plus de la thèse de la connaissance située et de la thèse de l'avantage épistémique, et pour mieux comprendre pourquoi la théorie féministe du *standpoint* n'est pas essentialiste, Sharon Crasnow (2013, 2018) ajoute un troisième point à la caractérisation de la théorie de la connaissance située développée par Intemann : la thèse de l'achèvement [*achievement thesis*]¹. Comme mentionné précédemment, les personnes en situation de subordination sociale ne connaissent pas automatiquement plus ou mieux (Wylie, 2003). Dans cette perspective, un *standpoint* et l'avantage épistémique lui étant associé se révèlent lorsqu'une conscience critique est exercée par rapport à la manière dont les structures de pouvoir limitent ou contraignent la production de connaissance. Autrement dit, le fait d'appartenir simplement à un groupe social subordonné n'est pas suffisant pour être considéré comme un *standpoint*. La thèse de l'achèvement a émergé à la suite de la critique selon laquelle la théorie de la connaissance située impliquerait une forme d'essentialisme (p. ex. : Hekman, 1997) et de l'interprétation selon laquelle l'avantage épistémique serait automatique (p. ex. : Pinnick *et al.*, 2003). Si le *standpoint* était automatique et essentialiste, le fait d'être une femme permettrait de connaître le monde selon la perspective « de la femme » de manière infaillible. Le fait d'être une femme donnerait en soi la capacité de mieux connaître certains aspects du monde. Or il est tout à fait possible que des individus aient internalisé leur oppression, ou que leur situation d'oppression les ait privés d'accès à des ressources éducatives permettant de comprendre les structures de pouvoir qui les affectent (Intemann, 2016). Ainsi, comme le défend Crasnow, pour pouvoir prétendre à un avantage épistémique, les individus appartenant à des groupes marginalisés doivent être conscients de leur place au sein des relations de pouvoir oppressives. L'avantage épistémique n'est pas donné ou automatique, mais atteint. Cela peut nécessiter, par exemple, de mettre en place des « groupes de

¹ La thèse de l'achèvement est présente chez Wylie (p. ex. : Wylie, 2003), mais elle n'est pas considérée comme un élément constitutif de la théorie du *standpoint*. Je suis d'accord avec Crasnow qu'en faire un élément constitutif de la théorie évite les confusions et les critiques d'essentialisme ou de privilège automatique qui ont été adressées à la théorie féministe du *standpoint*.

conscientisation » [*consciousness-raising groups*] (p. ex. : Anderson 2019; Crasnow, 2018; MacKinnon, 1989; Rolin, 2009a). Un exemple classique est celui de l'émergence du concept de harcèlement sexuel pour décrire des cas d'agressions subies par des femmes en milieu de travail et commises par leurs collègues masculins. L'expérience partagée des femmes en milieu de travail et l'analyse critique du phénomène d'un point de vue des relations de pouvoir en jeu ont permis de mettre au jour un phénomène d'oppression que n'aurait pas pu révéler l'expérience d'une seule femme (p. ex. : Intemann, 2016). Ce type de méthode permettrait que les individus appartenant à des groupes socialement subordonnés prennent conscience des facteurs sociaux et politiques qui maintiennent les situations d'oppression. Ainsi, pour plusieurs défenseurs de la théorie féministe du *standpoint*, l'idée de *standpoint* implique également la thèse de l'achèvement, s'éloignant par là d'un avantage épistémique automatique ou lié à une conception essentialiste.

5.2.2 Convergence de la théorie du *standpoint* et de l'empirisme féministe

Après avoir analysé les grandes lignes de la théorie féministe du *standpoint*, on peut à présent se demander si cette théorie est en compétition avec la théorie de l'OS, ou bien si elle n'en est qu'un complément. Disons d'emblée qu'à l'image de l'OS de Longino, la théorie féministe du *standpoint* soutient elle aussi une conception de l'objectivité sociale fondée sur la diversité. C'est ce que Sandra Harding a appelé « l'objectivité forte » (*strong objectivity*) (p. ex. : Harding, 2015). L'objectivité forte demande à ce que soient intégrés dans la communauté scientifique des membres occupant des *standpoints* marginalisés. L'objectivité sociale de la théorie féministe du *standpoint*

demande donc une diversité de *standpoints*, c'est-à-dire une diversité de membres appartenant à des positions sociales différentes. Dans cette perspective, les points de vue sociaux sont plus rigoureux et efficaces dans l'identification et la critique des biais et des présuppositions problématiques présents en science qu'une simple diversité de valeurs et d'intérêts. Cette efficacité accrue est soutenue par la thèse de l'avantage épistémique présentée dans la section précédente : les communautés en situation de subordination sociale sont plus à même d'identifier des schèmes méthodologiques et conceptuels émanant des communautés dominantes et d'en voir les limites. En effet, ces communautés ayant été traditionnellement exclues des pratiques scientifiques, elles peuvent apporter un regard neuf sur les paradigmes dominants opérant en science. L'intégration de leur point de vue maximise donc l'objectivité des communautés épistémiques.

Pour une raison similaire, Intemann (2010) soutient que la théorie féministe du *standpoint* ne devrait pas être conçue comme une alternative à l'empirisme féministe, mais bien comme un complément. Comme mentionné précédemment, dans l'empirisme féministe de Longino, la diversité de la communauté épistémique en est une de valeurs et d'intérêts. Plus une communauté est composée de membres dont les valeurs et intérêts varient, plus elle est socialement objective. Or cette injonction ne précise pas le type de valeurs et d'intérêts pertinents pour maximiser l'OS. Dans le cadre de la théorie féministe du *standpoint*, le type de diversité vient se préciser : une diversité de *standpoints*. Pour Intemann, la théorie de la connaissance située présente l'avantage de mettre en lumière les conséquences épistémiques d'une différenciation sociale sur ce que nous pouvons savoir comme agents épistémiques, et ainsi justifier l'adoption d'une diversité de *standpoints* au sein des communautés scientifiques. En cela, elle peut permettre de justifier l'inclusion de certains groupes sociaux marginalisés en science, au sens où les personnes appartenant à ces groupes (et qui seraient parvenues à une conscience critique de leur *standpoint*, conformément à la

thèse de l'achèvement) sont plus à même de déceler et de critiquer des éléments problématiques présents dans les présuppositions d'arrière-plan qui demeureraient invisibles pour les personnes appartenant à des groupes majoritaires ou privilégiés (Wylie, 2012).

Comme le note avec justesse Intemann (2016), la théorie féministe du *standpoint* est éminemment politique, beaucoup plus que l'OS de Longino. Elle vise à cartographier les différents types de relation de pouvoir et la manière dont elles affectent de manière différenciée la production de connaissances dans le but suivant : défaire ces relations de pouvoir inégalitaires dans l'intérêt des groupes marginalisés. L'objectif de la théorie féministe du *standpoint* est donc de produire une science dont les questions de recherche et les méthodologies participent à l'intérêt des groupes marginalisés. Dans cette optique, la méthodologie qui s'en suit prend comme point de départ l'expérience des personnes appartenant à des groupes socialement subordonnés pour produire des connaissances dans l'intérêt de ces groupes. Prendre comme point de départ le point de vue des groupes marginalisés permettrait la détection du point de vue dominant (valeurs, présuppositions, intérêts, etc.) (voir aussi Harding, 2015, p. 30). Intemann (2016) fait la liste des bénéfices épistémiques qui peuvent découler de l'adoption de la méthodologie de la théorie féministe du *standpoint*, conjointement à l'empirisme féministe :

- 1) L'identification et la remise en question des structures de pouvoir qui freinent la participation à la science des groupes marginalisés;
- 2) La correction et la minimisation des biais jugés problématiques (p. ex. : sexisme, racisme, etc.);
- 3) L'augmentation de la créativité et le pluralisme épistémique;
- 4) La création d'une science qui répond davantage aux besoins sociaux.

Ainsi, selon Intemann, la théorie de l'objectivité sociale de Longino devrait intégrer l'idée que les *standpoints* sont nécessaires à une réelle critique transformatrice intersubjective pour produire de réels bénéfices épistémiques dans certains contextes.

Le cadre de cette thèse ne me permet pas de me prononcer sur le bien-fondé de la proposition d'adjoindre à l'empirisme féministe la théorie de la connaissance située pour produire une théorie hybride dans tous les domaines d'investigation scientifique². D'ailleurs, Intemann soutient elle-même que la théorie hybride qu'elle propose ne vise pas à couvrir l'entièreté des domaines scientifiques, mais bien les domaines scientifiques dont les questions concernent ou affectent les groupes socialement subordonnés (Intemann, 2016, p. 276). Or dans le cas précis qui m'intéresse, celui de la définition du trouble mental dans le DSM, l'intérêt d'intégrer la perspective de personnes appartenant à des groupes marginalisés en psychiatrie me paraît convaincant, d'autant plus que ce type de diversité me semble compatible avec les objectifs poursuivis pour une communauté épistémique diversifiée. En effet, Longino plaide pour une diversité de points de vue et d'intérêts pour maximiser les chances d'identifier des présuppositions problématiques en science. La conception de la diversité de la théorie féministe du *standpoint*, avec la thèse de l'avantage épistémique, me semble fournir les ressources nécessaires pour justifier l'intégration de *standpoints* marginalisés au sein de la communauté. En effet, puisque la présence de personnes appartenant à des positions sociales subordonnées augmente les chances d'identifier des présuppositions ou des biais présents dans les cadres conceptuels dominants, ce type de diversité me paraît répondre aux objectifs de Longino. Ainsi, bien que Longino n'ait pas elle-même fait cette modification à sa théorie, je la crois compatible avec ses objectifs.

² La proposition d'Intemann ne fait pas consensus, voir Borgerson (2011) pour une critique.

Par ailleurs, rappelons que les femmes, les personnes homosexuelles, les patients.es et les groupes racisés/de culture non occidentale ont fait les frais de structures oppressives au sein de la psychiatrie (voir le chapitre 4 pour des exemples). Adopter la théorie féministe et empiriste du *standpoint* défendue par Intemann me semble particulièrement pertinent pour faire émerger des présuppositions d'arrière-plan qui seraient problématiques en psychiatrie et dans la définition du trouble mental. En ce sens, je vois peu de raisons de m'opposer à l'intégration de différents *standpoints* qui viendraient maximiser l'identification et la critique de biais en jeu dans la définition du trouble mental. Dans le chapitre 7, je soutiendrai que cette approche permet de fonder théoriquement l'inclusion de ces différents groupes au sein des structures de prises de décision du DSM, surtout par rapport à la définition du trouble mental.

5.3 L'objectivité sociale et l'intégration de communautés non scientifiques

La section précédente a montré que la manière dont Longino conçoit la diversité de la communauté épistémique pourrait être complétée par la théorie féministe du *standpoint*, du moins dans le contexte de la psychiatrie. Cette discussion a l'avantage d'attirer l'attention sur la composition de la communauté épistémique dans le modèle de l'OS que défend Longino et sur la notion même d'une communauté épistémique. Alors que la section précédente concernait le type de diversité requis à l'intérieur d'une communauté scientifique pour que celle-ci soit dite « socialement objective », cette section concerne les frontières de la communauté épistémique et les motifs qui justifient ces frontières. La notion de communauté épistémique est centrale dans l'approche de Longino, car c'est par la manière dont cette communauté est structurée

que l'on peut évaluer le degré selon lequel elle est socialement objective. Or la signification exacte d'une communauté épistémique est vague, et ses frontières précises, difficiles à identifier (p. ex. : Koskinen, 2017). Récemment, Alyson Wylie (2015) a défendu que bien que Longino conçoive l'OS comme s'appliquant uniquement à des communautés de scientifiques (ou de chercheur.e.s provenant du monde scientifique), la communauté épistémique devrait pouvoir être agrandie dans certains cas où l'inclusion de perspectives non scientifiques est pertinente. Dans cette section, je discuterai la pertinence d'ouvrir la communauté épistémique à des membres traditionnellement exclus de cette communauté, parce que non scientifiques. Cette discussion aura des implications pour le chapitre 7, où je défendrai que les patients devraient être intégrés aux prises de décision concernant la définition du trouble mental.

L'analyse de Wylie s'ancre dans la théorie féministe du *standpoint* présentée dans la section précédente. Rappelons que la méthodologie de la théorie féministe du *standpoint* demande de prendre comme point de départ le point de vue des individus appartenant à des groupes marginalisés ou subordonnés dans le but de mettre au jour les biais problématiques qui affectent la science. Argumentant pour une communauté épistémique plus inclusive, Wylie utilise des cas tirés de l'archéologie, où plusieurs projets de recherche ont été menés en collaboration avec des communautés autochtones (qui sont des communautés non scientifiques). Comme le montre Wylie, en archéologie, le fait d'être confronté à des traditions épistémiques différentes a permis une meilleure critique de la recherche et a fait aboutir de nouvelles connaissances. Lorsque les communautés autochtones étaient intégrées à la communauté épistémique, les normes méthodologiques et les présuppositions théoriques structurant les projets de recherches étaient plus facilement identifiables (et critiquables) que si la communauté épistémique était uniquement composée de membres provenant du milieu scientifique. Ces pratiques collaboratives entre communautés extrascientifiques et communautés

scientifiques auraient amené des transformations bénéfiques importantes en archéologie. L'intégration du point de vue de communautés extrascientifiques peut donc modifier la science positivement.

Un exemple utilisé par Wylie pour démontrer le caractère épistémiquement fructueux de l'interaction entre des communautés autochtones et des communautés de recherche en archéologie est l'apport de Roger Echo-Hawk, membre de la communauté autochtone des Pawnee³, située dans les plaines centrales des États-Unis (Wylie, 2015, p. 199-202). Formé comme historien et archéologue, Echo-Hawk s'est élevé contre un paradigme en archéologie qui consiste à percevoir le témoignage oral des communautés autochtones comme une source non valable d'éléments de preuves. Selon cette conception, la tradition orale est un matériel instable et peu fiable, voire la simple projection de croyances spirituelles autochtones. Echo-Hawk s'est opposé au rejet radical de la tradition orale comme objet d'étude utile pour reconstituer l'histoire des communautés autochtones. Echo-Hawk s'est penché sur les mouvements migratoires des communautés Pawnees du sud des États-Unis vers les plaines centrales, et ce, en se basant sur une analyse de la tradition orale de cette communauté. Par cette analyse, il a révélé que leur migration avait eu lieu beaucoup plus tôt que ce que la recherche en archéologie l'avait admis. Des preuves archéologiques ont en effet confirmé la présence des Pawnees sur les plaines centrales avant l'an 1000, alors que les historiens

³ La seule limite que je vois à l'exemple utilisé par Wylie est qu'Echo-Hawk s'est également formé comme historien et archéologue. Ainsi, à proprement parler, ce n'est pas « purement » un membre d'une communauté extrascientifique, puisque sa formation est de nature scientifique. Cependant, il est vrai que Hawk a importé dans ses recherches un système épistémique et des croyances propres à sa communauté autochtone. De ce fait, il s'agit en effet d'une interaction entre des systèmes épistémiques extrascientifique et intrascientifique, et ce, en dépit du fait que Hawk ait une formation d'historien et d'archéologue. À mon sens, bien que certaines précisions pourraient être apportées de la part de Wylie quant au statut particulier de Hawk (est-il réellement un membre d'une communauté extrascientifique?), sa formation ne me semble pas contredire l'argument de Wylie puisqu'il importe bel et bien dans ses travaux des éléments traditionnellement exclus de la communauté scientifique et qui sont propres à la communauté extrascientifique (autochtone), soit l'utilisation de la tradition orale pour guider les recherches archéologiques.

et archéologues avaient daté leur présence après 1600. Grâce à ce travail, Echo-Hawk a pu soutenir que la tradition orale peut offrir des « indices » de l'histoire des peuples autochtones, mais que son contenu demande à être confirmé par du matériel archéologique. Dans cette optique, la tradition orale autochtone et l'analyse archéologique deviennent complémentaires, et non pas opposées. Par ailleurs, la binarité entre la tradition orale des communautés autochtones, non valide et non scientifique, et la méthode archéologique, valide et scientifique, s'en trouve revue. Par cette étude de cas, Wylie soutient que l'interaction entre une communauté scientifique et une communauté extrascientifique (ici un membre de la communauté autochtone des Pawnees et une communauté scientifique d'archéologues et d'historiens) peut permettre la remise en question des normes épistémiques traditionnellement admises par rapport aux standards d'évidence acceptés (ici le témoignage oral des populations autochtones).

Pour rendre conforme l'intégration des communautés extrascientifiques au cadre conceptuel de l'OS, Wylie se base sur la quatrième norme de l'objectivité sociale, l'égalité tempérée d'autorité intellectuelle. Selon Wylie, rien dans la théorie de Longino ne justifie l'exclusion de communautés extrascientifiques qui seraient porteuses d'une expertise pertinente ou d'une perspective critique au regard d'une investigation scientifique particulière. En effet, l'OS de Longino demande à ce que soit entretenue une critique intersubjective transformatrice dans le but de faire émerger des biais et valeurs problématiques contenus dans les présuppositions d'arrière-plan. La pertinence d'intégrer des membres extrascientifiques à des projets de recherche s'explique par le fait que même si une communauté scientifique est diversifiée en fonction de différents groupes sociaux (p. ex. : femmes, groupes racisés, etc.), elle demeure uniforme quant à certains standards épistémiques communément acceptés. L'intégration d'acteurs qui ne proviennent pas du milieu scientifique, parce qu'elle

augmente la diversité des perspectives, rend alors plus faciles l'identification et la critique de ces normes (Wylie, 2015). Wylie écrit :

In order to counteract the risks of insularity and the effects of dysfunctional group dynamics that can insulate foundational assumptions and norms of justification from critical scrutiny, well-functioning epistemic communities should actively cultivate collaborations with external communities whose epistemic goals, practices, and beliefs differ from their own in ways that have the potential to mobilize transformative criticism. On this principle the impetus for dynamic, interactive pluralism is not just that it may fill lacunae and correct errors in the substantive beliefs of a research community, but that it can bring community members to a critical standpoint on their established knowledge making and ratifying practices (Wylie, 2015, p. 207).

Des membres de communautés extrascientifiques pourraient donc, dans certaines circonstances, être en mesure de mieux révéler les biais et valeurs problématiques que des membres internes à une communauté scientifique exclusive. Wylie défend que la norme de l'égalité tempérée d'autorité intellectuelle devrait s'appliquer à ces communautés extrascientifiques. Cette inclusion aurait pour effet que l'apport des communautés extrascientifiques à la science serait considéré et intégré, menant à ce qu'elle appelle un « pluralisme dynamique » de la recherche scientifique.

Avant de conclure cette section, il convient de noter que ce que suggère Wylie est compatible avec la théorie de l'OS de Longino (pour une discussion, voir Rolin, 2009b). En fait, Longino admet elle-même qu'une interaction entre des membres externes à une communauté scientifique et cette dernière peut être bénéfique pour déceler des présuppositions d'arrière-plan qui seraient admises par l'ensemble des membres de cette communauté (Longino, 2002, p. 133-134). Cependant, il est juste de dire que l'œuvre de Longino n'insiste pas vraiment sur cette interaction entre communautés scientifique et extrascientifique. Ainsi, en plus d'accepter l'intégration de différents *standpoints* pour maximiser la diversité de la communauté épistémique

(section 5.2.2), l'objectivité sociale devrait, pour les raisons invoquées par Wylie, permettre l'ouverture de la communauté scientifique à des membres non scientifiques, pourvu que cette ouverture permette la maximisation de l'identification des biais et des présuppositions d'arrière-plan. Cet amendement aura également des répercussions sur la manière de concevoir une définition objective du trouble mental, que je discuterai dans le chapitre 7.

5.4 Relativisme et objectivité sociale

Comme mentionné au début de ce chapitre, la théorie de l'objectivité sociale défendue par Longino a été abondamment discutée et critiquée. Jusqu'à maintenant, le développement de ce chapitre montre que la vaste majorité des discussions a concerné les aspects procéduraux et sociaux de cette conception de l'objectivité (les limites de la communauté épistémique, la manière d'opérationnaliser les quatre normes de l'OSP pour éviter l'antiféminisme, l'importance de la théorie féministe du *standpoint* pour une critique réellement transformatrice, etc.). Un aspect de ce modèle qui a été moins discuté dans les écrits est le lien entre l'objectivité sociale, l'empirisme et le relativisme. Pourtant, l'une des ambitions d'une théorie sociale de l'objectivité scientifique est de démontrer la *valeur épistémique* de l'accent mis sur les processus sociaux de production de connaissances. Plus précisément, une théorie de l'objectivité sociale comme celle défendue par Longino prétend offrir une « meilleure science » si les normes de l'OS sont mises en place, par contraste avec des modèles plus classiques de l'objectivité. Longino défend en effet qu'adopter son modèle de l'OS permet d'identifier des biais et présuppositions problématiques dans la recherche scientifique,

ce qui conduit à une amélioration de la science (p. ex. : Longino 1990, p. 215-220). Comme nous l'avons vu au chapitre 4, cette conception contraste avec l'idée traditionnelle de la science indépendante des valeurs pour laquelle c'est l'absence de valeurs et de préférences non épistémiques qui permet de maximiser l'objectivité scientifique, et ainsi donner un meilleur accès au monde tel qu'il est. Face à la vision classique de la science, la tâche de l'OS est alors de démontrer comment une science légitimement influencée par des valeurs non épistémiques pourrait mener à des meilleures connaissances sur le monde, ce qui peut apparaître contre-intuitif à première vue.

L'une des raisons qui expliquent la méfiance envers les sciences influencées par les valeurs est la présence de cas historiques extrêmement problématiques où l'adoption d'idéologies particulières pour guider la recherche scientifique a mené à une « mauvaise science ». Les sciences nazies sont un exemple paradigmatique d'une discipline où la présence de valeurs non épistémiques a clairement biaisé la recherche, voire où les valeurs non épistémiques ont remplacé les éléments de preuves. L'historien des sciences Robert Proctor (2000) a montré comment la science médicale nazie était perçue comme un simple outil devant servir les idéaux raciaux nazis (contre la thèse selon laquelle la médecine nazie était une pseudoscience). Comme le met bien en évidence Proctor, l'Allemagne était l'un des pays les plus avancés sur le plan médical, et plus généralement sur le plan scientifique, au moment de la Deuxième Guerre mondiale (Proctor cite en exemple les travaux très précis des chercheurs allemands sur le lien entre le tabac et l'amiante et le développement des cancers du poumon). Dans ce contexte, Proctor défend que la médecine nazie ne doit pas être perçue comme une pseudoscience, mais comme une véritable science médicale complice d'une « éthique » nazie (Proctor, 2000, p. 337). L'éthique médicale nazie idéalisait la pureté du corps et valorisait l'hygiène raciale. La médecine devait donc servir à promouvoir ces idéaux. À cet effet, on peut lire :

The Nazi medical profession, in other words, was not without its ethics. Nazi medical ethics was typically sexist and racist, but there was also stress on cleanliness, punctuality, orderliness, and obedience to legal authority, especially the supreme authority of Adolf Hitler, the beloved Führer. Nazi medical ethics tended to reduce morality to efficiency, economics, and aesthetics, relegating to the trash heap everything that was seen as ugly and a burden. Preventive medicine was emphasized, as was Nordic (and especially Germanic) supremacy, cost-efficiency, the natural lifestyle, and the superiority of the productive worker over persons deemed inferior or burdensome (Proctor, 2000, p. 343).

C'est en ce sens que l'on peut dire que les valeurs nazies orientaient la recherche médicale. L'infériorité des groupes « non blancs » était un postulat normatif, et non une découverte médicale. Les recherches médicales visaient à accumuler des connaissances et à développer des techniques pour pouvoir appliquer l'idéal racial nazi (menant aux horreurs que l'on connaît, soit le génocide de toutes les formes de vie humaines considérées comme inférieures). C'est en ce sens que Proctor écrit : « The primary failing of Nazi medicine, I would argue, was the failure of physicians to challenge the rotten, substantive core of Nazi values. » (Proctor, 2000, p. 344). La médecine nazie était pour Proctor une véritable science, influencée par des valeurs monstrueuses.

Face à l'analyse de Proctor de cet exemple radical, la question qui s'en suit est la suivante : si l'on accepte que la science puisse être influencée par les valeurs sans que cela affecte son statut épistémique et sa capacité de décrire le monde, disposons-nous de moyens pour juger l'exemple de la médecine nazie comme problématique? (voir par exemple Boghossian, 1996, p. 14 et Haack, 2003 pour ce type d'inquiétudes). En lien avec ces préoccupations, la critique qui me semble la plus problématique pour la théorie de l'OS est l'accusation de relativisme. Selon Crasnow (1993, 2003; voir aussi Freedman, 2009 et Smith, 2004 pour des critiques similaires), Longino remplacerait la réalité comme standard d'évaluation des théories et des hypothèses scientifiques par

une forme de consensualisme. Selon elle, le consensus (critique) au sein d'une communauté scientifique serait le seul critère pour justifier l'adoption de théories scientifiques particulières. L'OS impliquerait donc le relativisme parce que le lien entre les théories et le monde n'aurait pas de place dans le raisonnement scientifique. N'importe quelle hypothèse qui obtient consensus, si elle respecte les normes de l'OS, pourrait être qualifiée d'objective, ce qui apparaît insatisfaisant. La question de Crasnow est donc la suivante : l'objectivité sociale de Longino est-elle seulement sociale? Si oui, il semble que cette théorie ne permette pas d'éviter le relativisme. Le relativisme ayant mauvaise presse en philosophie des sciences, il convient de se pencher sur la manière dont Longino explore cette question et si sa théorie possède les ressources nécessaires pour éviter cette accusation.

5.4.1 Précisions sur la menace du relativisme

Avant de poursuivre, quelques clarifications s'imposent. La thèse du relativisme s'est présentée sous différentes formes qui ont été soutenues par de multiples arguments en philosophie (voir par exemple Baghramian et Carter, 2019; Boghossian, 2006; McAllister, 2017; Niiniluoto, 1996; Proctor, 1998 pour une revue des différentes acceptions et une discussion). La forme la plus problématique de relativisme est ce qui a été appelé le relativisme « radical ». Cette thèse implique que toutes les affirmations seraient d'égale valeur (Proctor, 1998). Si cette forme de relativisme est appliquée à la science, elle implique l'idée selon laquelle l'acceptation ou le rejet de toutes les théories et hypothèses scientifiques seraient également justifiés (ou injustifiés). Pour revenir au souci de Haack (2003) et de Boghossian (1996), selon cette interprétation

du relativisme, les idées de la médecine nazie seraient tout aussi justifiées que celles de la médecine contemporaine. J'interprète donc ces préoccupations comme relevant d'une accusation de relativisme radical (ou de menace de relativisme radical) adressée à l'objectivité sociale de Longino.

Bien que certains éléments de critique de Smith (2004), de Freedman (2009) et de Crasnow (1993, 2003) puissent viser le relativisme radical, ces attaques semblent viser avant tout une autre forme de relativisme, le « relativisme épistémique ». Le relativisme épistémique se définit comme le fait, pour chaque communauté épistémique, d'utiliser ses propres standards pour juger ses résultats/théories comme supérieurs. La justification épistémique est donc toujours relative à un système épistémique particulier. La relativité de la justification épistémique à des systèmes épistémiques particuliers rend difficile la comparaison des théories et des hypothèses scientifiques entre elles, étant donné qu'il existe plusieurs systèmes épistémiques différents et qu'il semble difficile de démontrer de manière non circulaire la supériorité d'un système épistémique par rapport à un autre (voir Baghramian et Carter, 2019 pour une discussion). Par exemple, dans le cas de l'OS de Longino, Crasnow (2003) défend que ce cadre théorique ne donne pas les moyens de penser la vérité scientifique au-delà des pratiques sociales et culturelles de justification épistémique. Autrement dit, la vérité scientifique serait relative, voire se réduirait aux préférences d'une communauté épistémique particulière. Cela mènerait Longino à adopter un relativisme épistémique. Les critiques de Smith et de Freedman sont moins précises, mais il me semble que si des solutions existent pour répondre aux menaces associées aux relativismes épistémique et radical, leurs critiques deviendront caduques.

Dans les deux prochaines sections, je vais donc m'intéresser particulièrement au relativisme en science et à ces deux types de relativisme, le relativisme radical et le relativisme épistémique, puisque ce sont ces deux formes qui semblent être en jeu dans

les critiques adressées à l'objectivité sociale de Longino sur le plan épistémique. Je vais soutenir que la théorie de Longino n'implique pas de relativisme radical, mais peut être interprétée comme un relativisme épistémique non problématique. Pour démontrer que la position de Longino ne conduit ni à un relativisme radical, ni à un relativisme épistémique problématique, il faut préciser comment sa théorie de l'OS s'ancre dans un empirisme contextualiste. Dans la section 5.4.2, je montrerai comment l'OS se rattache à l'empirisme. Dans la section 5.4.3, je discuterai la manière dont l'OS s'inscrit dans un contextualisme. Parce que la théorie de Longino est empiriste et contextualiste, elle ne mène ni à un relativisme radical, ni à un relativisme épistémique problématique.

5.4.2 Objectivité sociale et empirisme : l'adéquation empirique comme standard de justification nécessaire

Dans un premier temps, il convient d'approfondir en quoi l'OS se rattache à l'empirisme. Rappelons, comme exposé dans le chapitre 4, que l'expérience est le critère ultime de justification dans l'empirisme contextuel critique (ECC) de Longino. Comme le note avec justesse Bueter (2010), l'ECC de Longino combine des exigences sociales et empiriques pour penser l'acceptabilité épistémique (c'est-à-dire que l'acceptation d'une théorie ou d'une hypothèse est justifiée au regard des éléments de preuve pertinents). Pour bien comprendre comment l'OS est liée à cet empirisme, il est important de rappeler qu'elle demande le respect de quatre normes : 1) des lieux de critiques, 2) des standards d'évaluation publics, 3) une intégration de la critique et 4) une égalité d'autorité intellectuelle tempérée (voir le chapitre 4 pour l'exposé et la discussion de ces quatre normes). En lien avec la norme 2, en tant qu'empiriste,

Longino pose comme condition nécessaire d'un standard d'évaluation scientifique l'adéquation empirique entre les théories et les données (Longino, 1990, p. 77⁴, 1995). Ainsi, pour être acceptées au sein d'une communauté épistémique, les théories doivent se conformer aux données empiriques disponibles. Dans le cadre de l'OS, pour qu'une théorie soit acceptée, elle doit être conforme aux données empiriques et survivre à la critique intersubjective transformatrice. Ces deux conditions doivent être remplies. En ce sens, il est faux de dire que l'objectivité sociale conduit à abandonner les phénomènes comme juges des théories scientifiques. Longino dit plutôt que les phénomènes empiriques ne sont pas les seuls juges. Sa position n'implique donc pas un relativisme radical, puisque les théories et hypothèses doivent être adéquates sur le plan empirique.

Par ailleurs, il faut clarifier le rôle des valeurs non épistémiques au sein de l'ECC. Dire que les valeurs non épistémiques entrent en jeu dans le raisonnement scientifique ne signifie pas que les valeurs non épistémiques peuvent remplacer les éléments de preuve dans l'investigation scientifique, voire éliminer le besoin d'un recours aux éléments de preuve comme critère d'évaluation. Selon Longino, les données empiriques conservent leur rôle prépondérant dans la justification épistémique, comme dans le cas de la différenciation sexuelle (voir le chapitre 4). Les mesures du taux d'hormones et leur lien avec le développement de la morphologie sexuelle demeurent des éléments cruciaux de l'investigation scientifique. Cependant, la connexion entre ces données et les hypothèses est médiée par les présuppositions d'arrière-plan, où des valeurs épistémiques et non épistémiques peuvent être présentes (dans le cas de la différenciation sexuelle, il s'agit de l'hypothèse de la bicatégorisation sexuelle). Sur le fait que les données empiriques conservent un rôle prépondérant dans l'acceptabilité épistémique, Longino écrit :

⁴ Il convient toutefois de noter que dans cette discussion, Longino admet la possibilité d'interprétations divergentes de ce qui constitue l'adéquation empirique (Longino, 1990, p. 77; voir aussi Longino, 1995).

We are not forced to admit that anything goes. What we experience at a given time and place can be described and measured in conventionally fixed ways. Once we say of a mark on a gauge that it represents 10 units of a measure, we cannot read it as a 7 or a 13 without the consent of those with whom we need to communicate about the gauge and without changing other descriptions related to those of the gauge. It is neither our desires, nor beliefs, nor values, nor social conventions that make a gauge register a 10 when employed in measurement but processes independent of those natural phenomena. (Longino, 1990, p. 221-222)

Ainsi, bien que les valeurs puissent entrer dans les présuppositions d'arrière-plan, celles-ci ne remplacent pas les observations empiriques. Dans le cas de la mesure fournie par une jauge, une fois que tous les paramètres ont été fixés, c'est bien le monde qui dicte les données de la mesure, et pas les valeurs non épistémiques. Or, l'objectivité sociale permet de prendre en considération la présence possible de valeurs non épistémiques et donne des moyens pour identifier et critiquer ces valeurs. Ainsi, comme le soutenait Bueter (2010), deux types d'exigences sont bel et bien présents dans l'ECC : des normes concernant les processus sociaux de production de connaissances et l'appui de données empiriques pointant en faveur de l'adoption de théories ou d'hypothèses particulières.

Par ailleurs, pour bien contrer l'accusation de relativisme radical, il peut être utile de développer la position de Longino par rapport au réalisme scientifique étant donné que les débats sur le relativisme y sont souvent liés. Longino discute la thèse du réalisme scientifique dans son livre *Science as Social Knowledge* (1990, p. 28-32). Elle présente le réalisme scientifique comme renvoyant à l'idée selon laquelle les théories scientifiques sont vraies, c'est-à-dire qu'elles représentent le monde tel qu'il est (au moins approximativement)⁵. Par contraste, la thèse anti-réaliste empiriste suppose que

⁵ D'autres caractérisations du réalisme scientifique sont possibles. Par exemple, P. Godfrey-Smith (2003, p. 176) décrit le réalisme scientifique comme la thèse selon laquelle l'un des buts raisonnables de la science est de représenter le monde de manière fidèle (incluant les aspects de la réalité qui ne sont pas

la science cherche l'adéquation empirique entre les théories et les données pertinentes plutôt que la vérité. Par exemple, selon Van Fraassen (que Longino considère comme un anti-réaliste), le but de la science est de fournir des théories qui sont empiriquement adéquates (Van Fraassen, 1980, p. 12). Si les théories décrivent adéquatement les parties observables du monde, alors elles sont empiriquement adéquates. Si une théorie répond à ce critère d'adéquation empirique, alors il est raisonnable d'accepter cette théorie et d'user des concepts que propose la théorie pour résoudre des problèmes scientifiques. Les théories scientifiques ne viseraient donc pas à expliquer la structure de la réalité au-delà de ce qui est observable.

La position de Longino se situe à mi-chemin entre le réalisme et l'antiréalisme. Elle défend une forme de réalisme minimal, selon lequel une certaine réalité indépendante de l'esprit humain existe. Selon elle, il est possible de supposer que cette réalité viendrait imposer des contraintes à ce que l'on peut dire du monde à travers nos théories scientifiques (Longino, 2002, p. 222). Or, il ne serait pas possible de se prononcer sur la relation exacte qui unit nos théories à cette réalité. Les seules choses à la disposition des scientifiques sont les données empiriques. Le statut métaphysique de ces données n'est pas connaissable pour Longino, qui adopte par là une posture agnostique vis-à-vis le lien exact entre les théories et la réalité. Par ailleurs, pour elle, il serait faux de dire que le seul but de l'entreprise scientifique est de produire des énoncés vrais (au sens métaphysique) sur cette réalité. Elle écrit :

Whether the success of science demands that the world have the features attributed to it by successful theories depends on the aims of science. But there is not a single aim of science. There are multiple aims. A community of scientific realists will hold that one of the aims must be the true representation of the structure(s) of nature. A community of scientific antirealists or of

observables). Cette caractérisation ne se positionne pas sur le degré de fidélité de nos théories scientifiques actuelles à la réalité, mais sur les objectifs reliés à ces théories.

pragmatists will hold that empirically adequate or useful content is a sufficient aim. (Longino, 2002, p. 142)

Pour Longino, le lien exact entre les théories scientifiques et la structure de la réalité ne peut pas être connu, mais il est raisonnable de supposer qu'il existe.

Ainsi, il est faux de soutenir que l'OS mène à un relativisme radical, étant donné que Longino a développé sa théorie dans le cadre de l'empirisme et qu'elle admet un réalisme minimal. Il est donc également erroné de soutenir que l'OS implique seulement la régulation des processus sociaux de production des connaissances et que cette régulation conduit à diminuer l'importance de l'observation empirique comme justification à l'acceptabilité épistémique. Puisque l'OS s'ancre dans le cadre théorique de l'empirisme contextualiste, cette théorie est liée aux exigences de cet empirisme. Les théories et hypothèses scientifiques ne sont donc pas toutes aussi justifiées les unes que les autres, étant donné que les théories doivent se conformer aux données empiriques disponibles. Longino croit donc que l'on peut écarter certaines théories et hypothèses parce qu'elles ne se conformeraient pas aux données empiriques. Ainsi, l'OS s'appuie sur l'empirisme contextualiste critique, et évite par là le relativisme radical.

5.4.3. Contextualisme, objectivité sociale et relativisme épistémique

À la suite de la démonstration que la théorie de l'OS ne mène pas à un relativisme radical en science, il convient à présent de s'interroger sur l'accusation de relativisme épistémique. Dans les deux prochaines sections, je m'intéresserai particulièrement à la

critique formulée par Crasnow (1993, 2003) étant donné qu'elle me semble la plus développée dans les écrits sur cet enjeu. Je vais donc reconstituer la critique de Crasnow, dans le but de lui offrir la meilleure réponse possible. Comme je l'ai déjà mentionné, Crasnow (2003) soutient que l'OS de Longino conduirait à un relativisme épistémique, rendant incompatible l'objectivité sociale avec ce que serait une « bonne » philosophie des sciences. Par-là, Crasnow assume que le relativisme n'est pas une position soutenable en philosophie des sciences, qu'il s'agisse de relativisme radical ou épistémique (Crasnow, 2003, p. 139). Dans la section précédente, j'ai déjà argumenté que l'OS de Longino ne conduise pas à un relativisme radical. Dans cette section, je vais défendre qu'une certaine forme de relativisme épistémique, celle défendue par Longino, ne contredit pas une « bonne » philosophie des sciences, comme le soutient Crasnow.

Selon moi, c'est en approfondissant la manière dont le contextualisme de Longino se présente qu'il est possible de contrer l'accusation de relativisme épistémique. Rolin (2011) explore les différentes façons dont la théorie de Longino peut être qualifiée de contextualiste. Selon elle, le contextualisme peut se comprendre comme faisant référence⁶ 1) au contexte des présuppositions d'arrière-plan, 2) au contexte de la communauté épistémique ou 3) au contexte social et culturel de la science. Voyons à quoi ces trois formes de contextualisme réfèrent. Rappelons que pour Longino, les

⁶ Il faut noter que le contextualisme défendu par Longino (qui se retrouve de manière générale en philosophie féministe des sciences) est différent de ce qu'on appelle le contextualisme épistémique (bien que des points de contact existent). Evelyn Brister (2017) trace les similitudes et les différences entre les deux approches. Elle écrit : « While DeRose, Lewis, and Williams focus on relatively narrow issues concerning knowledge-attribution or epistemic justification, feminist epistemologists frequently emphasize the moral and political factors that both motivate epistemological inquiry and operate within particular contexts. » Cette précision tient aussi pour le contextualisme que l'on trouve chez Longino, qui est beaucoup plus large que le contextualisme épistémique. (Pour une tentative de faire correspondre une certaine forme de contextualisme épistémique avec le contextualisme de Longino, voir Bristler, 2017 et Rolin, 2011 pour une discussion; pour une explicitation du contextualisme épistémique en épistémologie classique, voir Rysiew, 2016.)

éléments de preuve qui permettent de justifier l'acceptation d'une théorie sont dépendants du contexte (voir le chapitre 4). La justification épistémique est en effet relative à un contexte de présuppositions d'arrière-plan qui participe à établir le lien de pertinence entre les éléments de preuve et les théories ou hypothèses. Les éléments de preuve sont donc toujours évalués relativement à un contexte particulier, étant donné que les scientifiques doivent s'appuyer sur des présuppositions d'arrière-plan méthodologiques et conceptuelles pour déterminer les éléments de preuve à considérer pour évaluer une hypothèse. Autrement dit, une information empirique est toujours considérée comme un élément de preuve en faveur ou en défaveur d'une théorie *au regard du contexte des présuppositions d'arrière-plan* (type 1) (Longino, 1990, p. 61). Ensuite, la communauté épistémique détermine les standards d'évaluation des théories et des hypothèses. Les normes qui régulent la justification épistémique (la quantité et la qualité d'éléments de preuve requis pour accepter une hypothèse, le choix entre des valeurs épistémiques à prioriser comme le pouvoir prédictif par rapport à la profondeur explicative, etc.) sont donc élaborées dans un contexte social particulier et ne peuvent être définies *a priori*. Les *standards d'évaluation épistémique sont donc toujours relatifs à un contexte particulier, celui de la communauté épistémique* au sein de laquelle l'investigation scientifique s'effectue (type 2) (Longino, 1990, p. 186). Pour finir, le contextualisme de type 3 renvoie à l'idée selon laquelle les valeurs du contexte social et culturel au sein duquel la science se pratique peuvent entrer dans le contexte des présuppositions d'arrière-plan et dans le contexte de la justification épistémique. La communauté scientifique pratique ses recherches dans un contexte social et culturel particulier, qui peut influencer le contexte des présuppositions d'arrière-plan (1) ou le contexte qu'est la communauté épistémique (2). Selon moi, pour simplifier, il est possible de dire que le type 3 peut être intégré aux types 1 et 2. En effet, le contexte de type 3 intervient dans le contexte 1 : les valeurs non épistémiques propres au contexte social et culturel de l'investigation scientifique peuvent intercéder dans le contexte des présuppositions d'arrière-plan qui interviennent dans la justification épistémique. Il

peut aussi intervenir dans le contexte de type 2, puisque le contexte social et culturel peut venir déterminer les standards épistémiques à prioriser. Dans la suite de la discussion, j'assumerai donc qu'il y a deux formes de contextualisme chez Longino : le contexte des présuppositions d'arrière-plan (1) et le contexte de la communauté épistémique (2). Il est utile de distinguer les deux, parce que ces deux types de contextualisme impliquent des problèmes distincts et nécessitent des solutions différentes.

Dans un premier temps, l'OS fournit des outils pour lutter contre un relativisme épistémique problématique entraîné par le contextualisme de type 1. L'élément fondamental de l'OS qui constitue la réponse de Longino est le suivant : l'investigation se produit au sein d'une communauté, et c'est cette communauté qui, par sa nature critique, vient réguler les autres facteurs qui affectent l'acceptation ou le rejet d'une théorie. En faisant en sorte que le contexte des présuppositions d'arrière-plan soit explicité et donc possiblement critiquable grâce aux quatre normes de l'objectivité sociale, ce ne sont pas tous les éléments qui peuvent être admis dans les présuppositions d'arrière-plan. Seules les présuppositions qui auront survécu à la critique pourront demeurer dans le contexte pour la suite de l'investigation. Il est donc vrai de dire que les éléments de preuve sont relatifs au contexte des présuppositions d'arrière-plan, et donc qu'ils sont dépendants du contexte. Or puisque les présuppositions d'arrière-plan sont critiquables et doivent survivre à la critique pour persister, il est faux de dire que toutes les présuppositions d'arrière-plan se valent. Cette forme de contextualisme n'implique pas de relativisme épistémique problématique. Comme Crasnow le reconnaît, l'objectivité sociale se présente bel et bien comme une solution à ce problème. Or, éviter ce problème grâce à l'objectivité sociale n'implique pas qu'un autre problème est résolu, soit celui amené par le contextualisme de type 2 (la relativité des standards de justification par rapport au contexte d'une communauté épistémique particulière).

Rappelons que le relativisme épistémique concerne la relativité de l'évaluation des théories et des hypothèses au contexte de la communauté épistémique/du système épistémique utilisé. Boghossian résume ainsi le problème du relativisme épistémique : « there are no universal epistemic facts, that facts about what belief is justified by a given item of evidence can vary from community to community [...] different people may rationally arrive at opposed conclusions, even as they acknowledge all the same data. » (Boghossian, 2006, p. 59). L'une des raisons pour lesquelles le contextualisme de Longino peut être interprété comme impliquant un relativisme épistémique problématique (où tous les systèmes épistémiques seraient équivalents) est qu'il est possible d'imaginer que dans certains contextes, une justification épistémique satisfaisante d'un énoncé ou d'une hypothèse ne dépende pas de son soutien empirique. Par exemple, un contexte épistémique pourrait être celui d'une communauté religieuse, où la foi et l'interprétation des textes religieux constituent les seuls éléments de preuve nécessaires à l'acceptation d'une théorie (voir Brister, 2017). Ainsi, si la justification épistémique est relative au contexte épistémique, et si cette communauté épistémique est une communauté religieuse, alors on pourrait penser qu'il est légitime d'accepter des affirmations comme « Dieu est parfait » ou « La théorie de l'évolution est fausse ». En effet, étant donné que, pour cette communauté, les standards de justification impliquent uniquement la foi et une interprétation des textes religieux, et que ceux-ci pointent effectivement vers la vérité d'affirmations telles que « Dieu est parfait », cette affirmation pourrait être considérée comme « prouvée », ce qui contrevient à nos intuitions sur ce qu'est une « bonne » science. Cet exemple fait écho à la critique de Crasnow (2003), pour qui des expériences mystiques pourraient être jugées comme des éléments de preuve en faveur de l'existence de Dieu dans certaines communautés épistémiques (voir la section 5.4). Bien qu'il soit possible que certaines formes de contextualisme impliquent ce type de cas, imputer ce problème au contextualisme de Longino est, comme on le verra, une erreur.

En effet, penser ainsi serait considérer son contextualisme comme indépendant de son empirisme. Or les deux vont de pair, ce qui conduit Longino à poser que le critère de l'adéquation empirique est un critère nécessaire (mais non suffisant) pour que soit acceptée une théorie ou une hypothèse, comme exposé dans la section 5.4.2. L'empirisme de Longino et l'accent mis sur l'adéquation empirique impliquent des observations et des expérimentations systématiques et possiblement répétables (*reproducible*). Comme la foi et l'expérience mystique ne répondent pas aux exigences de l'empirisme, on peut exclure ces pratiques de la science (pour une discussion, voir Erciyes, 2015, p. 190). Ainsi, parce que le contextualisme de Longino est *empiriste*, il ne peut être soumis à une telle accusation. Bien que plusieurs critères de justification et d'évaluation des théories et des hypothèses puissent intervenir de façon différente dépendamment des communautés épistémiques (p. ex. : la simplicité des théories peut être valorisée dans certains contextes, alors que le caractère explicatif d'une théorie le sera dans un autre), le seul critère qui doit être partagé *par toutes les communautés épistémiques* pour qu'elles puissent être qualifiées de « scientifiques » est qu'elles recherchent l'adéquation empirique. Pour Longino, ce critère n'est donc pas contextuel, mais plutôt définitionnel de ce qu'est une « bonne » science. Grâce à l'empirisme, il est possible de hiérarchiser des systèmes épistémiques (p. ex. : un système épistémique reposant sur la foi est moins bon qu'un système épistémique reposant sur l'observation et l'expérimentation).

Avant de conclure cette section, il convient de noter, comme le fait Intemann (2016), que même le critère d'adéquation empirique a quelque chose de contextuel. Elle écrit :

Empirical adequacy depends on additional shared standards and assumptions about the cognitive aims of inquiry. Some of these aims and standards will be social, ethical, and political value judgments about what features of the world we take to be salient or important to project. These assumptions will be relevant to determining whether a model or theory accounts for the relevant data or is able to make empirical predictions of the right kind (Intemann, 2016, p. 267).

En ce sens, l'adéquation empirique est un critère nécessaire à l'acceptation ou au rejet des théories scientifiques, mais la manière exacte selon laquelle cette notion est comprise par la communauté épistémique peut varier. Or un lien entre les éléments de preuve et les théories scientifiques doit toujours être présent, bien que la nature de ce lien puisse varier (conformément au réalisme scientifique minimal que défend Longino, voir la section 5.4.2). En ce sens, le relativisme épistémique de Longino n'est pas réellement problématique, en ce qu'il ne conduit pas aux types d'exemples décrits par Crasnow et Bristler, qui relèveraient d'un relativisme épistémique *radical, sans critère empirique*. Pour Longino, tous les systèmes épistémiques ne se valent pas. Seuls ceux qui adoptent le critère de l'adéquation empirique peuvent être qualifiés de « scientifiques ». C'est donc en ancrant sa théorie de l'objectivité sociale dans le cadre conceptuel de l'empirisme contextualiste que Longino parvient à éviter un relativisme épistémique problématique. En effet, pour Longino, les standards de justification des théories sont bel et bien relatifs à un contexte particulier, et en ce sens peuvent varier, mis à part le critère déterminant pour qu'une communauté scientifique soit dite « appartenir à la science » : l'adéquation empirique.

Pour résumer, l'imbrication de l'objectivité sociale, du contextualisme et de l'empirisme permet à Longino de ne pas sombrer dans un relativisme radical, tout en admettant la présence de valeurs sociales et culturelles au sein de l'investigation scientifique et en reconnaissant le caractère contextuel de l'évaluation épistémique.

5.5 Analyse de la signification de l'objectivité scientifique : objectivité métaphysique, objectivité empirique et objectivité méthodologique

Un enjeu en lien avec celui du relativisme, mais qui n'est pas équivalent, est la signification de l'objectivité scientifique et son rapport avec l'objectivité sociale. Bien qu'il apparaisse désormais clair que l'objectivité sociale, parce qu'elle s'ancre dans le cadre conceptuel de l'empirisme contextualiste critique, ne remplace pas l'observation empirique par le consensus (et évite par là le relativisme radical et une forme problématique de relativisme épistémique), il est tout de même vrai de dire que l'objectivité sociale réfère aux processus sociaux de production de connaissance. Il convient maintenant de se pencher sur les différents sens du mot « objectivité scientifique » et sur son lien avec l'objectivité sociale. Cette section concerne donc davantage l'analyse du concept d'objectivité scientifique plutôt que la menace de relativisme entraînée par l'objectivité sociale. L'enjeu ici tourne autour de la pluralité des sens du terme « objectivité scientifique ». Il s'agira de déterminer si l'objectivité sociale telle que décrite par Longino se présente comme suffisante au regard de l'analyse conceptuelle de l'objectivité scientifique. Pour Crasnow (1993, 2003), l'accent que met Longino sur les processus sociaux de production de savoir scientifique est problématique, puisqu'une conception de l'objectivité scientifique implique à la fois une référence aux procédures de production de savoir et à *la vérité scientifique* (Crasnow, 2003, p. 134). Parce que le concept d'objectivité sociale ne ferait pas intervenir le concept de « vérité », il serait insatisfaisant pour rendre compte de nos intuitions sur ce qu'est l'objectivité scientifique.

Pour répondre à cette inquiétude, il faut revenir à la manière dont Longino conçoit la signification de l'objectivité scientifique. Selon cette dernière, traditionnellement, le

concept d'objectivité a eu plusieurs significations. Selon elle, deux significations principales peuvent être distinguées : 1) l'objectivité métaphysique/ontologique et 2) l'objectivité méthodologique⁷. L'objectivité métaphysique implique que les hypothèses et théories scientifiques représentent le monde tel qu'il est (Longino, 1990, p. 62). Ce sens du mot objectivité dénote la relation entre les théories scientifiques et le monde. Une théorie objective décrit ou explique les objets du monde et les relations qu'ils entretiennent entre eux (p. ex. : causalité, probabilité, etc.), conformément à ce qu'ils sont « réellement ». Par contraste, l'objectivité méthodologique réfère à des processus ou à des critères non arbitraires et non subjectifs pour évaluer les hypothèses scientifiques. Selon Longino, l'objectivité méthodologique peut référer à deux choses. Dans un premier temps, l'objectivité méthodologique fait référence à la fiabilité des résultats scientifiques suivant des méthodes appropriées de collecte et d'interprétation de données. Dans un deuxième temps, l'objectivité méthodologique indique plutôt le caractère « public » de la science (*publicity of science*) (Longino, 1990, p. 70). Ce type d'objectivité que Longino appelle « publique » réfère à trois exigences : 1) un langage partagé pour décrire et investiguer les phénomènes à l'étude, 2) les phénomènes à l'étude étant perçus comme existant indépendamment de nous et 3) l'examen des biais et croyances idiosyncrasiques dans l'évaluation des théories au regard des éléments de preuves. Le troisième aspect de la science est pour Longino ce qui lui confère réellement l'objectivité. Dans cette perspective, un accord justifié de manière critique et intersubjective est ce qui confère une valeur épistémique à une théorie ou à une hypothèse scientifique. En ce sens, si des processus non arbitraires et non subjectifs, tels que les normes de l'OS, sont utilisés pour accepter ou rejeter une théorie scientifique, alors ce mode d'investigation mis en place par une communauté épistémique peut être dit « objectif ».

⁷ Cette caractérisation de l'objectivité selon deux sens principaux, ontologique et méthodologique, est partagée par plusieurs auteurs (tels que Datson, 1992; Megill, 1994; voir Erciyes, 2015 pour une discussion). Or je rappelle que d'autres caractérisations sont possibles (voir le chapitre 4).

Selon Longino, la « sagesse populaire » défend que l'objectivité méthodologique est ce qui permet de garantir l'objectivité métaphysique (Longino, 1990, p. 63). En cherchant une meilleure objectivité méthodologique, il devient possible d'espérer une meilleure objectivité métaphysique, c'est-à-dire un « meilleur accès au monde ». C'est donc l'objectivité méthodologique qui intéresse au premier chef Longino. Cependant, comme le note Ercyes (2015, p. 247), sur le plan conceptuel, l'idée de méthode implique un but à atteindre. Dans ce cas-ci, le but d'une méthode objective est de parvenir à une meilleure description du monde. Or, la théorie de Longino se concentre explicitement sur l'objectivité méthodologique et ne s'embarrasse pas d'expliquer le lien entre objectivité méthodologique et objectivité métaphysique. Cela est d'autant plus problématique qu'il n'est pas clair que l'objectivité méthodologique assure l'objectivité métaphysique. La critique de Crasnow me semble donc pointer un élément important et possiblement problématique de la théorie de Longino : son insistance sur les processus sociaux d'investigation scientifique (objectivité méthodologique) écarte un autre sens que nous attribuons généralement à l'objectivité scientifique, soit la relation entre la connaissance et le monde (objectivité métaphysique). L'exigence selon laquelle ce qui est décrit par les théories scientifiques existe bel et bien dans la nature peut apparaître comme un élément central du concept d'objectivité scientifique, élément auquel la science doit son autorité épistémique. Longino devrait donc montrer comment l'objectivité méthodologique, sous la forme de l'objectivité sociale, entraîne une meilleure description du monde, sans adopter la vision selon laquelle il est possible d'accéder au monde indépendamment de nous (sous un angle différent de la « perspective de nulle part » (*view from nowhere*) ou « la vue de Dieu » (*God eye's view*), que rejette Longino).

À la lumière de l'analyse conceptuelle qu'offre Longino de l'objectivité scientifique, il convient désormais de clarifier un autre point sémantique de la théorie de l'OS de Longino. Comme je l'ai montré dans la section précédente, dans l'ECC, la présence

d'éléments de preuve est indispensable pour que soit acceptée une théorie ou une hypothèse, conformément au critère nécessaire d'adéquation empirique. Or pour Longino, la présence d'éléments de preuve empiriques en soutien à une théorie ou à une hypothèse scientifique n'est pas ce qui confère à ces théories ou hypothèses leur *objectivité*, mais plutôt ce qui leur confère leur *scientificité*. Dans ce cadre, l'objectivité est atteinte par les modes de production de connaissances scientifiques, et ne réfère pas au produit de ce processus. Longino écrit :

The respect in which science is objective, on this view, is one that it shares with other modes of inquiry, disciplines such as literary or art criticism and philosophy. The feature that has often been appealed to as the source of the objectivity of science, *that its hypotheses and theories are accepted or rejected on the basis of observational, experimental data, is a feature that makes scientific inquiry empirical.* [...] In the contextual analysis of evidential relations, however, that a method is empirical in the above sense does not mean that it is also objective. (Longino, 1990, p. 75, les italiques sont les miennes)

Ainsi, pour Longino, les données empiriques sont ce qui donne le caractère scientifique à la science, mais pas ce qui rend la science objective. L'objectivité sociale est donc une étiquette que peut porter d'autres domaines de production de savoir, comme l'éthique, la politique ou l'esthétique. Il est donc juste de noter que pour Longino, ce sont les processus sociaux de production de connaissances qui, s'ils sont organisés pour permettre la CIT, garantissent une meilleure objectivité. C'est le caractère empirique des objets d'études de la science qui lui donne sa scientificité. Étant donné que l'objectivité sociale « scientifique » doit toujours être comprise en relation avec l'empirisme contextualiste de Longino, comme je l'ai montré dans la section précédente, l'enjeu est moins de savoir si la relation avec les données empiriques est exclue de cette conception, mais plutôt si la conceptualisation qu'offre Longino de l'objectivité est satisfaisante. La question est donc de savoir si l'analyse conceptuelle de l'objectivité scientifique que présente Longino est convaincante.

Une approche intéressante qui peut aider à éclairer cette question est celle de la philosophe de l'épistémologie sociale Marianne Janack (2002). Comme le note la chercheuse, le concept d'objectivité est un idéal. Or, ce qui complique les discussions entourant ce concept est que plusieurs idéaux lui sont rattachés. Janack identifie en effet 14 significations différentes impliquant toutes des idéaux différents et affirme qu'il y en a probablement plus. En lien avec le problème qui nous occupe, relevons qu'elle affirme que l'idéal des procédures objectives de production du savoir et l'idéal métaphysique de vérité sont deux idéaux différents qui ne vont pas nécessairement ensemble. On peut atteindre l'idéal de procédures objectives sans pour autant atteindre l'idéal métaphysique (Janack, 2002, p. 278; voir aussi Douglas, 2004, 2009). Ce que Janack suggère est que, pour penser l'objectivité scientifique, il faut réfléchir aux rôles que joue ce concept et aux idéaux épistémiques auxquels il réfère dans des contextes particuliers d'investigation scientifique, plutôt qu'à sa signification purement abstraite (pour un argument similaire, voir Brister, 2017 et Koskinen, 2018).

Il faut donc réfléchir au rôle que joue l'OS dans la théorie de Longino. Étant donné que l'un des objectifs principaux de Longino est de maintenir la légitimité de la science par rapport aux autres disciplines tout en permettant une vision de la science incluant les valeurs non épistémiques, l'absence du sens métaphysique de l'objectivité peut paraître un enjeu de taille. Le lien entre l'objectivité métaphysique et l'objectivité méthodologique est un enjeu vaste et complexe. Mon but dans cette section est simplement de suggérer une piste de solution pour rendre plus satisfaisante la conception de l'objectivité scientifique défendue par Longino, afin de montrer que ce type de critique n'est pas nécessairement fatal pour sa théorie. Je suggère qu'il est possible d'élaborer une notion « d'objectivité empirique », alternative à l'objectivité métaphysique, qui serait compatible avec le réalisme minimal de Longino. Cette alternative pourrait répondre à l'insatisfaction des critiques visant l'incomplétude de la notion de l'OS pour rendre compte de l'objectivité scientifique. Bien qu'il soit vrai que

l'OS se concentre principalement sur l'aspect procédural de l'objectivité, Longino défend un réalisme minimal. Celui-ci implique l'idée d'une relation minimale entre les théories et le monde, mais exclut la thèse selon laquelle il serait possible d'accéder (et de savoir que l'on accède) au « réellement réel » (*[Really Real]*, selon l'expression de Lloyd, 1995). En ce sens, une objectivité métaphysique qui impliquerait un accès au monde naturel tel qu'il est « absolument » est incompatible avec l'approche de Longino. Elle écrit :

To say that a theory or hypothesis was accepted on the basis of objective methods does not entitle us to say it is true but rather that it reflects the critically achieved consensus of the scientific community. In the absence of some form of privileged access to transempirical (unobservable) phenomena it's not clear we should hope for anything better (Longino, 1990, p. 79).

Or, ce n'est pas parce que Longino rejette la thèse d'un accès au monde « tel qu'il est » qu'elle est conduite à proscrire toute référence au monde empirique. En effet, une référence au monde empirique est nécessaire pour qu'une théorie scientifique soit acceptée, conformément au critère d'adéquation empirique prôné par Longino (voir la section précédente). Or la nature exacte du lien entre les théories et le monde n'est pas connue a priori, ce pour quoi Longino demeure agnostique à ce sujet. Néanmoins, le fait que l'objectivité méthodologique (l'OS) implique nécessairement l'adéquation empirique comme critère la conduit à soutenir un réalisme minimal, au sens où les théories doivent pouvoir se rapporter minimalement au monde (mais pour Longino, la relation entre ce que nous savons du monde « tel qu'il est » grâce à nos théories et aux autres représentations du monde et le monde « tel qu'il est » est impossible à connaître, parce que nous ne pouvons pas dépasser nos outils de connaissance).

Le fait que Longino argue pour la nécessité des données empiriques dans l'évaluation des théories scientifiques me semble correspondre à ce que l'on pourrait appeler une « objectivité empirique ». L'« objectivité empirique » pourrait rendre compte du lien

entre l'objectivité métaphysique et l'objectivité méthodologique (l'objectivité empirique reste de l'ordre de l'ontologique, mais minimalement, étant donné que Longino défend un réalisme minimal, où le lien exact entretenu entre les données empiriques et le monde ne peut pas être connu). Pour répondre à la critique du manque d'objectivité de l'OS, je suggère donc que si une notion d'objectivité empirique était intégrée dans la théorie de l'objectivité scientifique de Longino, elle rendrait cette conception plus satisfaisante, intégrant à la fois les processus sociaux qui mènent à la production de connaissances scientifiques, c'est-à-dire l'objectivité méthodologique, et une notion minimale d'objectivité ontologique (de relation entre la connaissance et le monde), c'est-à-dire l'objectivité empirique. Ces deux sens de l'objectivité sont essentiels pour que la valeur et l'autorité qui est généralement attribuée à la science soient préservées, et donc pour que l'OS soit satisfaisante au regard de ses ambitions de parvenir à démontrer la possibilité d'une science influencée par les valeurs qui soit légitime. Dans le chapitre suivant, je vais explorer plus avant la manière dont pourrait s'élaborer cette notion d'objectivité empirique, et comment l'objectivité empirique pourrait prendre forme dans le cadre d'une définition scientifiquement objective du trouble mental.

Conclusion

Dans ce chapitre, je me suis penchée sur les principales critiques qui ont visé la notion d'objectivité sociale. J'ai d'abord abordé les critiques qui ont visé les aspects sociaux de l'OS, puis ses aspects plus épistémiques. Dans les deux prochains chapitres, je vais

montrer comment l'OS, sous la forme défendue par Longino, et les amendements acceptés dans cette section peuvent être appliqués au concept de trouble mental.

Références du chapitre 5

- Anderson, E. (2019). Feminist epistemology and philosophy of science. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/sum2019/entries/feminism-epistemology/>
- Baghramian, M. et Carter, J. A. (2019). Relativism. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/win2019/entries/relativism/>
- Boghossian, P. (1996, 13 décembre). What the Sokal hoax ought to teach us. *Times Literary Supplement*, 14-15.
- Boghossian, P. (2006). *Fear of knowledge : Against relativism and constructivism*. New York : Clarendon Press.
- Borgerson, K. (2011). Amending and defending critical contextual empiricism. *European Journal for Philosophy of Science*, 1(3), 435-449.
- Briester, E. (2017). Feminism and contextualism. Dans J. J. Ichikawa (dir.), *The Routledge Handbook of Epistemic Contextualism* (p. 57-67). New York : Routledge.
- Briester, E. (2018). Objectivity in science : The impact of feminist accounts. Dans P. Garavaso, *The Bloomsbury companion to analytic feminism* (p. 212-235). Londres : Bloomsbury Academic.
- Büter, A. (2010). Social objectivity and the problem of local epistemologies. *Analyse & Kritik*, 32(2), 213-230.
- Clough, S. (2003). *Beyond epistemology : A pragmatist approach to feminist science studies*. Landham : Rowman and Littlefield.

- Collins, P. H. (1986). Learning from the outsider within : The sociological significance of Black feminist thought. *Social Problems*, 33, 14-32.
- Crasnow, S. L. (1993). Can science be objective? Longino's *Science as social knowledge*. *Hypatia*, 8(3), 194-201.
- Crasnow, S. L. (2003). Can science be objective? Dans C. Pinnick, N. Koertge et R. Almeder (dir.), *Scrutinizing feminist epistemology : An examination of gender in science* (p. 130-141). Piscataway, NJ : Rutgers University Press.
- Crasnow, S. (2006). Feminist anthropology and sociology : Issues for social science. Dans S. Turner et M. Risjord (dir.), *Handbook of the philosophy of science, volume 15 : Philosophy of anthropology and sociology* (p. 755-789). Amsterdam : Elsevier.
- Crasnow, S. (2013). Feminist philosophy of science : Values and objectivity. *Philosophy Compass*, 8(4), 413-423.
- Crasnow, S. (2018). Contemporary standpoint theory : Tensions, integrations, and extensions. Dans P. Garavaso (dir.), *The Bloomsbury companion to analytic feminism* (p. 188-211). Londres : Bloomsbury Publishing.
- Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, 138, 453-473.
- Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh : University of Pittsburgh Press.
- Erciyes, B. (2015). *Situating objectivity : A feminist conceptualization* (thèse de doctorat). York University, Toronto.
- Freedman, K. L. (2009). Diversity and the fate of objectivity. *Social Epistemology*, 23(1), 45-56.
- Haack, S. (2003). Knowledge and propaganda. Dans C. Pinnick, N. Koertge et R. Almeder (dir.), *Scrutinizing feminist epistemology : An examination of gender in science* (p. 7-19). Piscataway, NJ : Rutgers University Press.

- Hankinson-Nelson, L. (1990). *Who knows? From Quine to a feminist empiricism*. Philadelphia : Temple University Press.
- Harding, S. G. (1986). *The science question in feminism*. Ithaca, NY : Cornell University Press.
- Harding, S. (2015). *Objectivity and diversity : Another logic of scientific research*. Chicago : University of Chicago Press.
- Hekman, S. (1997). Truth and method : Feminist standpoint theory revisited. *Signs : Journal of Women in Culture and Society*, 22(2), 341-365.
- Hicks, D. (2011). Is Longino's conception of objectivity feminist? *Hypatia*, 26(2), 333-351.
- Intemann, K. (2008). Increasing the number of feminist scientists : Why feminist aims are not served by the underdetermination thesis. *Science & Education*, 17(10), 1065-1079.
- Intemann, K. (2010). 25 years of feminist empiricism and standpoint theory : Where are we now? *Hypatia*, 25(4), 778-796.
- Intemann, K. (2011). Diversity and dissent in science : Does democracy always serve feminist aims? Dans H. E. Grasswick (dir.), *Feminist epistemology and philosophy of science* (p. 111-132). Dordrecht : Springer.
- Intemann, K. (2016). Feminist standpoint. Dans L. Disch et M. Hawkesworth (dir.), *The Oxford handbook of feminist theory* (p. 261-282). Oxford : Oxford University Press.
- Janack, M. (2002). Dilemmas of objectivity. *Social Epistemology*, 16(3), 267-281.
- Koskinen, I. (2017). Where is the epistemic community? On democratisation of science and social accounts of objectivity. *Synthese*, 194(12), 4671-4686.

- Koskinen, I. (2018). Defending a risk account of scientific objectivity. *The British Journal for the Philosophy of Science*.
- Lloyd, E. A. (1995). Objectivity and the double standard for feminist epistemologies. *Synthese*, 104(3), 351-381.
- Longino, H. (1993). Subjects, power, and knowledge : Description and prescription in feminist philosophies of science. Dans L. Alcof et E. Potter (dir.), *Feminist epistemologies* (p. 101-120). New York : Routledge.
- Longino, H. E. (1995). Gender, politics, and the theoretical virtues. *Synthese*, 104(3), 383-397.
- MacKinnon, C. A. (1989). *Toward a feminist theory of the state*. Cambridge, MA : Harvard University Press.
- Mcallister, J. W. (2017). Relativism. Dans W. H. Newton-Smith (dir.), *A Companion to the philosophy of science* (p. 405-407). Oxford : Blackwell Publishers.
- Niiniluoto, I. (1996). The relativism question in feminist epistemology. Dans L. H. Nelson et J. Nelson (dir.), *Feminism, science, and the philosophy of science* (p. 139-157). Dordrecht : Springer.
- Pinnick, C. L., Koertge, N. et Almeder, R. F. (dir.). (2003). *Scrutinizing feminist epistemology : An examination of gender in science*. Piscataway, NJ : Rutgers University Press.
- Proctor, J. D. (1998). The social construction of nature : Relativist accusations, pragmatist and critical realist responses. *Annals of the Association of American Geographers*, 88(3), 352-376.
- Proctor, R. (2000). Nazi science and Nazi medical ethics : Some myths and misconceptions. *Perspectives in Biology and Medicine*, 43(3), 335-346.
- Reiss, J. et Sprenger, J. (2017). Scientific Objectivity. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/>

- Rolin, K. (2009a). Standpoint theory as a methodology for the study of power relations. *Hypatia*, 24(4), 218-226.
- Rolin, K. (2009b). Scientific knowledge : A stakeholder theory. Dans J. Van Bouwel (dir.), *The social sciences and democracy* (p. 62-80). Basingstoke : Palgrave Macmillan.
- Rolin, K. (2011). Contextualism in feminist epistemology and philosophy of science. Dans H. E. Grasswick (dir.), *Feminist epistemology and philosophy of science* (p. 25-44). Dordrecht : Springer.
- Rolin, K. (2017). Can social diversity be best incorporated into science by adopting the social value management ideal? Dans K. C Elliott et D. Steel (dir.), *Current controversies in values and science* (p. 113-129). New York : Routledge.
- Rysiew, P. (2016). Epistemic Contextualism. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/win2016/entries/contextualism-epistemology/>
- Smith, T. (2004). “Social” objectivity and the objectivity of values. Dans P. Machamer et G. Wolters (dir.), *Science, values, and objectivity* (p. 143-171). Pittsburgh : University of Pittsburgh Press.
- Solomon, M. (2007). *Social empiricism*. Cambridge, MA : MIT Press.
- Solomon, M. (2006). Norms of epistemic diversity. *Episteme*, 3(1-2), 23-36.
- Steel, D., Fazelpour, S., Gillette, K., Crewe, B. et Burgess, M. (2018). Multiple diversity concepts and their ethical-epistemic implications. *European Journal for Philosophy of Science*, 8(3), 761-780.
- Wylie, A. (2003). Why standpoint matters. Dans R. Figuerroa et S. Harding (dir.), *Science and other cultures : Issues in philosophies of science and technology*. New York : Routledge.

Wylie, A. (2012). Feminist philosophy of science : Standpoint matters. *Proceedings and Addresses of the American Philosophical Association*, 86(2), 47-76.

Wylie, A. (2015). A plurality of pluralisms : Collaborative practice in archaeology. Dans F. Padovani, A. Richardson et J. Y. Tsou (dir.), *Objectivity in science : New perspectives from science and technology studies*. (p. 189-210). Dordrecht : Springer.

CHAPITRE 6

OBJECTIVITÉ EMPIRIQUE ET THÉORIE DE LA FONCTION COMME RÔLE CAUSAL

Introduction

Il est maintenant temps d'appliquer la théorie sociale de Longino, telle qu'elle a été amendée dans le chapitre 5, au cas du trouble mental. La difficulté à laquelle on fait face lorsque vient le temps d'appliquer la notion d'objectivité sociale au cas du trouble mental réside dans le caractère normatif particulier de ce phénomène. Le problème majeur de ce chapitre est celui-ci : bien que le trouble mental soit considéré en psychiatrie comme un objet de connaissance empirique, ses frontières semblent uniquement dessinées par les valeurs sociales et culturelles (épanouissement personnel, bien-être, absence de souffrance, conception de la vie bonne, ou valeurs problématiques telles que racisme, sexisme, hétéronormativité, capacitisme, etc. [pour des illustrations des différentes valeurs en jeu, voir les chapitres 2, 3 et 4]). En effet, suite aux conclusions des chapitres 2 et 3, les conceptions objectivistes du trouble mental paraissent insatisfaisantes. L'option normativiste semble donc la plus prometteuse, bien qu'elle demande à être amendée avec les outils de l'OS. Cependant, conformément aux conclusions du chapitre 5 selon lesquelles l'objectivité scientifique doit être à la fois sociale et empirique, le défi est de taille : il faut montrer que le concept

de trouble mental peut être objectif sur le plan de la critique intersubjective, *mais aussi* qu'un contenu empirique objectif relatif au concept de trouble mental peut être obtenu (pour répondre aux desiderata d'objectivité sociale *et* d'objectivité empirique). La question devient alors la suivante : comment prétendre que le concept de trouble mental est objectivement empirique si ce qui est constitutif de la définition du trouble mental semble relever davantage des valeurs sociales et culturelles que des données empiriques?

La raison pour laquelle cette situation est exceptionnelle est que la vaste majorité des écrits qui se sont intéressés aux rôles des valeurs sociales et culturelles en science a utilisé des exemples où les données empiriques demeuraient les premiers constituants des théories scientifiques, des hypothèses et des concepts. Les valeurs non épistémiques venaient ensuite compléter les données. Par exemple, l'ECC de Longino s'intéresse à des cas d'études où les données empiriques constituent le socle des théories scientifiques (comme les études sur le dimorphisme sexuel, où les données empiriques sont de l'ordre de la corrélation entre les hormones sexuelles et la morphologie et le comportement en fonction du sexe. Les valeurs non épistémiques viennent ensuite expliquer la catégorisation binaire « homme/femme », mais ne remplacent pas les données [voir le chapitre 4]). Il y a donc une zone d'ombre en philosophie des sciences et en épistémologie sociale quant à des concepts comme celui de trouble mental : un concept dont les composantes *majeures* sont des valeurs sociales et culturelles, mais qui ont également une composante empirique.

Dans le but de combler cet écart, je vais me pencher à nouveau sur les théories de la fonction et leur lien avec le concept de trouble mental. Je propose de revisiter les théories de la fonction pour voir si certains éléments pourraient être compatibles avec une vision de la science légitimement influencée par les valeurs. Je vais suggérer que le langage fonctionnel, caractéristique de la NFO, peut être réinterprété à l'aune de la théorie de l'objectivité sociale, si une théorie « perspectivaliste » de la fonction est

adoptée en me basant sur les travaux de Craver (2001, 2005, 2007, 2013). La théorie de la fonction perspectivaliste admet l'existence de fonctions qui reposeraient sur des mécanismes causaux, mais admet que l'identification de ces fonctions dépendrait ultimement de nos intérêts humains. Moyennant certaines modifications, le concept de trouble mental pourra être dit socialement *et* empiriquement objectif, et donc scientifique, grâce à une théorie de la fonction perspectivaliste. Dans ce qui suit, je vais donc rappeler certaines grandes lignes des théories de la fonction, puis je vais présenter un argument en faveur d'une théorie de la fonction perspectivaliste complétée par les normes de l'objectivité sociale.

6.1. Les théories de la fonction comme rôle causal et le concept de trouble mental

Étant donné que pour qu'un concept de trouble mental soit objectif, il doit respecter les normes de l'objectivité sociale *et* être associé à du contenu empirique, il me semble pertinent de revenir aux intuitions de l'approche objectiviste des fonctions naturelles (OFN). En effet, l'OFN plaide pour que la définition du trouble mental intègre une composante empirique. Rappelons que l'OFN a vu dans la notion de dysfonction une possibilité d'ancrer le concept de trouble mental dans le domaine de l'empirique et de l'objectif. Comme je l'ai présenté dans les chapitres 2, 3 et 4, l'approche OFN a conceptualisé la notion de dysfonction comme indépendante des valeurs sociales et culturelles, dans le cadre de l'idéal de la science indépendante des valeurs. Ayant remis en question cet idéal et ayant défendu que de bonnes pratiques scientifiques pouvaient être influencées par les valeurs si elles étaient régulées par les normes de l'objectivité sociale (chapitre 4), il convient désormais de revisiter la notion de dysfonction dans le

cadre de l'objectivité sociale. Dans les chapitres 2 et 3, j'ai montré comment les théories de la fonction « orientée vers les buts » (OB) de Boorse et de l'effet sélectionné (ES) de Wakefield n'étaient pas satisfaisantes pour définir le concept de trouble mental. Les critiques que j'ai présentées ciblaient, entre autres, le fait que ces deux approches ne parvenaient pas à atteindre leurs ambitions d'indépendance vis-à-vis les valeurs. Or s'il est possible qu'une bonne science soit légitimement influencée par les valeurs, il devient également possible de conserver la notion de dysfonction dans une définition du trouble mental, même si elle est influencée par les valeurs. La notion de dysfonction pourrait être complétée par les normes de l'objectivité sociale pour être satisfaisante. Dans cette section, je propose donc de revisiter le débat sur la normativité des théories de la fonction et leur lien avec la définition du trouble mental, et ce, avec les outils de l'objectivité sociale. Cet examen me servira à lier une théorie de la fonction et de la dysfonction perspectivalistes au concept de trouble mental pour parvenir à un concept objectif du point de vue social et empirique.

D'abord, revenons aux motivations de l'approche OFN qui expliquent le recours à la notion de dysfonction. L'idée était de trouver un critère normatif permettant de départager le normal du pathologique, ce critère devant provenir des sciences naturelles, et non du social ou du politique, considérés comme arbitraires (p. ex. : la notion de dysfonction devait pouvoir départager des cas controversés comme l'homosexualité ou l'alcoolisme sans recourir aux valeurs, jugées instables et variables, McDougall, 2020). Se tourner vers une théorie de la fonction est donc apparu aux défenseurs de l'OFN comme la solution au problème de la normativité du trouble mental, étant donné que pour la grande majorité des approches, une théorie de la fonction a pour ambition de fournir une norme *naturelle* permettant de départager la fonction de la dysfonction. En effet, comme le défend Wouters (2005, p. 127), l'un des enjeux centraux du débat sur les fonctions en philosophie a été celui de la normativité. Une théorie de la fonction *normale* devait permettre d'établir ce qu'un organe *doit* réaliser comme activité et de quelle manière cet organe *doit* la réaliser. Dans une théorie

de la fonction normale, ce type d'information peut être obtenu même si l'organe ne réalise pas sa fonction ou s'il ne la réalise pas de la « bonne » manière. On peut alors parler d'organe dysfonctionnel puisqu'il ne réalise pas l'activité qu'il devrait réaliser. Une théorie de la fonction permet donc d'identifier quelle est la norme de fonctionnement pour un système particulier (p. ex. : une théorie de la fonction normale devrait permettre d'établir que la fonction du cœur est de pomper le sang, selon un niveau d'efficacité particulier : le cœur est donc « supposé » pomper le sang, avec un niveau d'efficacité X, pour être dit fonctionnel, c'est sa fonction normale). De façon similaire, si le concept de fonction normale est appliqué à la psychiatrie, une théorie de la fonction devrait permettre de déterminer quels traits et comportements mentaux sont fonctionnels ou dysfonctionnels (Garson, 2006, p. 82). C'est avec ces notions de fonctions normales et de dysfonction qu'il serait possible de distinguer le normal du pathologique (p. ex. : s'il est établi que l'une des fonctions normales de l'esprit est la régulation des émotions, l'anxiété chronique et la dépression majeure peuvent être conçues comme des dysfonctions, et donc comme pathologiques¹). Qui plus est, il est généralement posé que cette norme de fonctionnement doit être descriptive et factuelle (Garson, 2006, p. 83) pour éviter les dérives associées à la surmédicalisation et au contrôle social par la psychiatrie (voir l'introduction). Recourir à une théorie de la fonction normale permet donc de générer une normativité essentielle pour départager le normal du pathologique sans faire appel à la normativité éthique, sociale, culturelle, esthétique, etc. Les écrits sur les fonctions ont donc semblé particulièrement attrayants en philosophie de la médecine et de la psychiatrie pour définir le trouble mental (pour une interprétation similaire, voir Roe et Murphy, 2011).

¹ Rappelons que pour Wakefield, il faudrait aussi que cette dysfonction cause un préjudice pour être un trouble mental (voir le chapitre 3), mais la dysfonction est tout de même un élément nécessaire du concept de trouble mental.

Bien qu'une théorie de la fonction normale ait été perçue comme le moyen de fonder la distinction entre le normal et le pathologique sur une normativité biologique, les théories de la fonction qui ont été utilisées jusqu'à maintenant n'atteignent pas cette ambition. Cela veut-il dire qu'il faille abandonner tout langage fonctionnel en psychiatrie? Selon moi, ce que les critiques ont montré est plutôt que les valeurs jouent un rôle crucial dans le langage fonctionnel en médecine et en psychiatrie, et non pas que le langage fonctionnel doit être rejeté. Il me semble donc pertinent de repenser les théories de la fonction à la lumière de cette conclusion. En m'appuyant principalement sur les travaux de Craver (2001, 2005, 2007, 2013) ainsi que ceux de Roe et Murphy (2011), je souhaite soutenir qu'une théorie de la fonction qui a généralement été écartée du débat sur la définition du trouble mental par son caractère trop libéral puisse retrouver sa pertinence, si amendée. Cette conception, défendue à l'origine par Cummins, est appelée *théorie de la fonction comme rôle causal* (RC). Dans le cadre épistémologique développé dans cette thèse, la RC pourrait retrouver ses titres de noblesse en psychiatrie. Cette légitimation de la RC sera rendue possible en appliquant les principes de l'objectivité sociale. Dans cette optique, ce seront les normes de l'objectivité sociale qui viendront donner les contraintes aux attributions fonctionnelles dont la RC avait besoin. Qui plus est, avec les amendements apportés récemment à la RC par Craver et Roe et Murphy, c'est le contenu empirique (mécanismes causaux et idéalizations) des attributions fonctionnelles qui viendra donner au trouble mental son caractère scientifique, et donc objectivement empirique. Avant de poursuivre, notons que mon but dans ce chapitre n'est pas d'offrir une revue complète et exhaustive des diverses théories de la fonction (voir Garson, 2016 et Wouters, 2005 pour d'excellentes présentations). Étant donné que j'ai déjà présenté les grandes lignes de l'OB et de l'ES dans les chapitres précédents, je vais circonscrire mon analyse à la manière dont la théorie RC et ses versions amendées pourraient donner du contenu empirique à la définition du trouble mental. Je vais donc laisser de côté les aspects de cette théorie de la fonction qui s'écartent de ce point focal, ainsi que d'autres aspects du débat sur les théories de la fonction qui ne sont pas rattachés à ce fil conducteur.

6.1.1 La théorie de la fonction RC

Une théorie qui a été peu discutée dans le débat sur la définition du trouble mental est l'approche de la fonction comme rôle causal (RC), développée par Robert Cummins (1975, 1983; une exception en philosophie de la médecine est Lange, 2007). Cummins a élaboré une théorie de la fonction ayant pour but de rendre compte de la logique explicative qui sous-tend les attributions fonctionnelles en physiologie et en psychologie (voir aussi Amundson et Lauder, 1994; Craver 2001, 2007, 2013; Davies, 2001; Hardcastle, 1999, 2002; Roe et Murphy, 2011 pour de plus récentes élaborations ou défenses). Pour Cummins, certaines pratiques en science ne font pas appel à des explications historiques, et donc évolutionnistes, lorsqu'elles attribuent des fonctions à des traits ou à des comportements. Cummins présente donc une alternative à l'ES en affirmant que les sciences sont également intéressées à expliquer le fonctionnement actuel des systèmes qui les intéressent, et plus particulièrement certaines capacités de ces systèmes (Cummins, 1975, p. 386-387). Cummins vise donc à rendre compte des attributions fonctionnelles à des *capacités* (p. ex. : reproduction) dans des systèmes complexes (p. ex. : les organismes) sans référence à leur histoire sélective passée. En ce sens, une fonction est la contribution causale d'une composante à la capacité générale d'un système duquel elle fait partie. Ainsi, si une observatrice attribue une fonction dans un système particulier, cela équivaut à attribuer un rôle causal à une composante du système qui participe à produire la capacité générale du système. L'attribution fonctionnelle ne vise donc pas les raisons pour lesquelles cette capacité est présente au sein d'un système (comme le ferait l'ES en invoquant l'histoire évolutionniste du trait ou de la capacité), mais plutôt le rôle causal d'une composante participant à la réalisation de la capacité d'un système (Cummins, 1975, p. 762). Selon Cummins, pour expliquer la réalisation d'une capacité associée à un système, les scientifiques décomposent généralement le système en sous-systèmes. Le découpage

est fait pour rendre compte du rôle causal de chacun des sous-systèmes à la production de la capacité du système général choisie pour l'étude. La fonction d'un sous-système est donc son activité/action qui permet la réalisation de la capacité générale à expliquer de manière causale. Pour plus de simplicité, reprenons l'exemple classique de la fonction du cœur. Un observateur pourrait affirmer que la fonction du cœur est de pomper le sang s'il s'intéresse à la circulation sanguine au sein de l'organisme humain. Le cœur, en tant que composante du système « circulation sanguine », dont la capacité est de circuler le sang au sein de l'organisme humain, participe causalement à cette capacité en pompant le sang. Pour la RC, la fonction du cœur est donc de pomper le sang.

Il est à noter que pour Cummins, les attributions fonctionnelles doivent se comprendre comme une stratégie explicative relative aux intérêts de l'investigateur. Ainsi, par contraste avec la théorie de Boorse, les fonctions à l'étude ne sont pas nécessairement les activités qui participent à l'atteinte des buts vitaux de l'organisme que seraient la survie et la reproduction. Les fonctions à l'étude sont plutôt les capacités d'un système qui intéressent les scientifiques (qui peuvent être la survie et la reproduction, mais pas uniquement). À l'inverse de l'approche ES, l'effet de la sélection passée ne vient pas contraindre les attributions fonctionnelles. Il n'y a donc pas de contraintes pour guider les attributions fonctionnelles, mis à part les intérêts des scientifiques et la manière dont les composantes participent causalement à la réalisation d'une capacité d'un système (Amundson et Lauder, 1994, p. 448). Ainsi, c'est parce que la circulation sanguine intéresse les scientifiques que pomper le sang est considéré comme la fonction du cœur. C'est l'activité qui explique causalement le rôle du cœur dans la circulation sanguine : c'est donc sa fonction. Il est possible d'imaginer que dans un autre cadre explicatif et des intérêts de recherche divergents, le cœur se voie attribuer d'autres fonctions. Or n'est-il pas vrai de dire que dans la pratique scientifique, il semble que la fonction du cœur soit bel et bien de pomper le sang? Pour Cummins, certaines explications sont intéressantes, et d'autres non, ce qui explique pourquoi certains traits et comportements

ne sont pas considérés comme des fonctions. Si le cœur ne se voit pas attribuer d'autres fonctions, c'est que le cœur ne fait pas partie d'autres systèmes qui intéressent actuellement les scientifiques².

Deux arguments sont généralement avancés en faveur de la RC. D'abord, la RC semble mieux décrire la manière dont les attributions fonctionnelles prennent place en dehors de la biologie et de la psychologie évolutionnistes (p. ex. : Amundson et Lauder, 1994; Roe et Murphy, 2011). Ensuite, la CR ne requiert pas qu'un trait ou un comportement ait été sélectionné pour qu'ils puissent se voir attribuer une fonction, ni que les fonctions se rapportent à des buts vitaux de l'organisme, promettant par là de dépasser les objections adressées à l'OB et à l'ES comme exposées dans les chapitres 2 et 3. Malgré ces bénéfices par rapport aux autres approches de la fonction, la RC n'a généralement pas été retenue en philosophie de la médecine et en philosophie de la psychiatrie. Dans la section suivante, j'exposerai les raisons qui ont motivé ce rejet.

6.1.2 Critiques de l'approche RC

Un problème majeur a généralement été pointé par les critiques de l'approche RC : sa trop grande libéralité (p. ex. : Kitcher, 1993; Millikan, 1989;). Cette trop grande libéralité aurait deux effets jugés problématiques (Garson, 2016, p. 85-89) : 1) la RC ne pourrait pas distinguer une fonction normale d'un accident, en ce qu'elle permettrait d'attribuer des fonctions à des traits qui sont généralement considérés comme *non*

² Or Cummins demeure vague par rapport à ce qui permet de distinguer, pour un investigateur, ce qui est intéressant de ce qui ne l'est pas (comme le note Craver, 2001, p. 57).

fonctionnels; 2) la RC n'offrirait pas le type de normativité attendu d'une théorie de la fonction, en ce qu'elle ne permettrait pas d'identifier des traits et comportements comme *dysfonctionnels*.

D'abord, abordons la première limite de la RC reliée à sa libéralité, soit son incapacité à distinguer la fonction de « l'accident », c'est-à-dire un bénéfice corrélé qui n'est pas une fonction. Garson (2016) appelle ce problème celui des traits « non fonctionnels » (1). Puisque dans la RC, l'attribution de fonction est relative aux intérêts des scientifiques, il est possible de penser qu'un même trait ou comportement pourrait se voir attribuer plusieurs fonctions, selon les intérêts qui guident cette attribution. La RC ne donne pas les moyens de discriminer ou de prioriser certaines attributions fonctionnelles par rapport à d'autres, tant et aussi longtemps que ces attributions sont conformes aux intérêts des scientifiques³. Une illustration de cette relativité des attributions fonctionnelles considérée comme problématique est l'incapacité de la RC à distinguer *la* fonction du cœur, qui est de pomper le sang, d'accidents ou d'effets corrélés à cette fonction, tels que le fait d'émettre un battement sonore lorsque le sang est pompé. En effet, supposons qu'un investigateur s'intéresse plutôt au bruit produit par le cœur et à son rôle pour détecter des possibles problèmes cardiaques lorsque les médecins écoutent le battement sonore du cœur de leurs patients à l'aide d'un stéthoscope. Le système à l'étude pourrait alors être décomposé différemment : le système devient la relation patient-médecin, et le cœur devient alors une composante

³ Il est à noter que Cummins identifie certaines contraintes à l'attribution fonctionnelle à des systèmes à l'étude (Cummins, 1975, p. 191). Pour Cummins, les attributions fonctionnelles sont appropriées lorsque : 1) les sous-capacités sont moins sophistiquées que la capacité à l'étude; 2) les sous-capacités sont de nature différente de la capacité à l'étude; 3) l'organisation des sous-capacités entre elles est relativement complexe. Ces critères sont plutôt vagues, et Cummins ne précise pas vraiment comment ils s'appliquent à des capacités particulières. En ce qui nous concerne, ces contraintes ne sont pas liées à la normativité des attributions fonctionnelles; elles ne visent donc pas à résoudre le problème de la dysfonction et des traits non fonctionnels. Il est cependant possible de dire que ce sont les intérêts des investigateurs qui déterminent si une attribution fonctionnelle est à prioriser par rapport à une autre, mais ces intérêts sont bien la seule contrainte.

qui joue un rôle causal diagnostique dans ce système grâce au battement sonore qu'il produit. Dans ce système, on s'intéresse à la capacité diagnostique du médecin, et le cœur joue bel et bien un rôle causal par le battement sonore qu'il produit (voir par exemple Garson, 2016, p. 83). Dans ce contexte d'investigation, la fonction du cœur est d'émettre un battement sonore et non de pomper le sang, parce que le bruit contribue causalement à la détection de problèmes cardiaques, à la capacité à expliquer, au sein du système « patient-médecin ». Selon cette critique, la RC n'aurait pas les ressources, du moins dans la formulation présentée par Cummins, pour exclure le bruit du cœur des attributions fonctionnelles considérées comme légitimes. Précisons cependant que la RC a les ressources pour départager les attributions fonctionnelles une fois les intérêts des investigateurs rendus explicites. Si l'intérêt est la capacité « circulation sanguine », il n'est pas correct de défendre que la fonction du cœur est d'émettre un battement sonore. Dans ce cadre, la seule fonction du cœur est de pomper le sang étant donné que le bruit émis par le cœur ne participe pas causalement à la circulation sanguine. Mais si l'intérêt des investigateurs est la capacité « diagnostic émis par un médecin », le battement sonore du cœur peut être sa fonction (voir Garson, 2016, p. 86 pour la clarification). La RC permet donc l'attribution fonctionnelle à des traits ou à des comportements que l'on juge généralement non fonctionnels, parce que cette attribution se fait relativement à des intérêts ou à des perspectives⁴.

Un deuxième problème est que la RC ne peut pas offrir la normativité attendue d'une théorie de la fonction. Garson appelle cette limite le problème des traits « dysfonctionnels » (2) (Garson, 2016, p. 87)⁵. Ce problème se manifeste de deux

⁴ Il est à noter que cette critique se base sur une intuition considérée comme communément partagée, celle que la fonction du cœur est de pomper le sang. Pour les auteurs qui endossent cette critique, l'idée ici est qu'il faut pouvoir rendre compte de cette intuition de manière satisfaisante. Une solution possible serait d'argumenter que cette intuition est infondée, ou n'est pas communément partagée.

⁵ Cette critique ne fait pas l'unanimité, voir Godfrey-Smith 1993 et Craver 2001 pour des positions divergentes. Ce qui est clair, c'est que la RC ne peut pas fournir une distinction entre le fonctionnel et le dysfonctionnel qui serait indépendante des intérêts humains, ce qui est perçu par Garson comme un

façons : 2.A) le manque de ressources pour attribuer un statut dysfonctionnel à un trait qui ne réalise pas sa « fonction » (ou ce qui est généralement considéré comme sa fonction) et 2.B) l'attribution fonctionnelle à des traits ou à des comportements généralement considérés dysfonctionnels. D'abord, examinons 2.A). Supposons à nouveau qu'un observateur s'intéresse à la circulation du sang, mais que l'organisme à l'étude est mort (et donc ne possède pas un cœur qui pompe le sang) ou que l'organisme présente un problème cardiaque particulier (le cœur pompe le sang, mais selon un niveau d'efficacité insuffisant). Dans les deux cas, la RC ne pourra pas considérer ce cœur comme dysfonctionnel. Dans le cadre de la RC, si le cœur ne pompe pas le sang, le cœur ne participe pas causalement à la capacité à l'étude, la circulation sanguine. Ainsi, le cœur n'aura tout simplement pas de fonction. Si le cœur pompe le sang selon un niveau d'efficacité insatisfaisant, il contribue tout de même causalement à la circulation du sang, et donc ne peut pas être dit dysfonctionnel. La CR ne peut donc pas rendre compte d'états généralement considérés comme dysfonctionnels, ce qui apparaît problématique. Dans un deuxième temps (2.B), la RC permet de considérer des traits intuitivement dysfonctionnels en tant que fonctionnels. Cela s'explique encore une fois par le fait que l'attribution d'une fonction à un trait ou à une composante d'un système est relative aux intérêts des investigateurs. Ainsi, si des investigateurs s'intéressent à des pathologies, il serait possible d'identifier des sous-systèmes qui expliquent le développement d'une pathologie comme « fonctionnels ». Si le cancer ou la dépression sont ce qui intéresse les scientifiques, alors les mutations cancéreuses ou la diminution de la transmission de la sérotonine pourraient être dites « fonctionnelles ». En effet, ces mécanismes causaux, parce qu'ils participent causalement à maintenir le cancer ou la dépression, les capacités à l'étude, pourraient être identifiés comme des fonctions « normales ». Selon Garson (2016), ce résultat apparaît contre-intuitif, puisque le cancer et la dépression représentent des exemples

problème.

paradigmatiques de dysfonctions (ou du moins de pathologies, auxquelles on n'associe généralement pas le qualificatif de « fonctionnelles », voir aussi Nervi, 2010).

Ce type de critique a généralement émergé dans les écrits sur les fonctions en biologie. Or, je crois qu'elles doivent être prises en compte en ce qui concerne la distinction entre le normal et le pathologique en psychiatrie. Il est en effet possible d'en tirer des implications pour une définition du trouble mental qui voudrait faire usage de la RC. Ce que montrent les critiques présentées précédemment, c'est que la RC n'est pas capable de rendre compte de la distinction entre le fonctionnel, le dysfonctionnel ou le non fonctionnel de manière satisfaisante. C'est sa trop grande libéralité liée à la relativité des attributions fonctionnelles aux intérêts humains qui explique pourquoi l'approche RC n'a pas été utilisée dans les définitions de l'approche objectiviste des fonctions naturelles. Étant donné que ce qui était recherché par l'OFN était les ressources normatives et *naturelles* des théories de la fonction, la relativité des attributions fonctionnelles aux intérêts humains (des scientifiques) a paru comme une contradiction avec le desideratum d'objectivité d'une théorie de la fonction normale. En effet, étant donné les deux problèmes caractéristiques de la RC présentés plus haut, son usage apparaît problématique en psychiatrie. Comment pourrait-on déterminer si l'homosexualité, le deuil ou le trouble dysphorique prémenstruel sont des dysfonctions dans ce cadre, si la seule contrainte aux attributions fonctionnelles réside dans les intérêts humains? N'y a-t-il pas un risque que des conditions comme l'homosexualité soient considérées comme une pathologie si les intérêts de chercheurs sont influencés par l'homophobie? Le problème est le suivant : en fonction des intérêts humains, chacun des cas énumérés pourrait être dit fonctionnel ou dysfonctionnel. La RC ne pourrait pas à elle seule départager les cas controversés en psychiatrie, puisqu'elle ne fournit pas de critères « indépendants » pour démarquer le normal du pathologique. Elle échouerait donc à une des tâches premières d'une définition des troubles mentaux, soit la production d'un critère de démarcation entre les conditions pathologiques et normales.

Dans ce qui suit, je vais présenter deux propositions d'amendement qui s'inscrivent dans la lignée de la RC, tout en ayant pour but de dépasser les objections précédemment mentionnées. Pour chacune de ces propositions, je vais examiner si elles restreignent de manière satisfaisante l'approche de Cummins pour que cette théorie de la fonction puisse faire partie d'une définition du trouble mental de manière satisfaisante.

6.2 La solution de Craver : attributions fonctionnelles, mécanismes causaux et perspectivalisme

Récemment, Carl Craver a proposé de sauver la théorie de la RC en joignant les notions de mécanisme et de fonction dans un cadre pragmatiste, défendant ce qu'il appelle *l'approche mécanique du rôle causal* (acronyme MRC, *mechanistic causal role*, Craver, 2001, 2005, 2007, 2013; voir aussi Amundson et Lauder, 1994; Hardcastle, 1999, 2002; Schaffner, 1993; Wimsatt, 1972 pour d'autres approches pragmatistes de la fonction). Cette approche tente d'ajouter des contraintes à la RC et ainsi de restreindre sa trop grande libéralité par l'ajout de la notion de mécanisme (Craver, 2001). De façon générale, Craver s'est intéressé aux explications mécanistes en neurosciences ainsi qu'au rôle des attributions fonctionnelles dans ces explications. Pour Craver, les théories en neurosciences décrivent des mécanismes, et les attributions fonctionnelles sont une façon d'intégrer ces différents mécanismes dans un réseau hiérarchique cohérent (Craver, 2005).

Pour bien comprendre la MRC, il faut d'abord préciser ce que Craver entend par « mécanisme », étant donné que cette notion est elle-même sujet de débat en philosophie des sciences (p. ex. : Bechtel et Richardson, 1993/2010; Craver, 2001;

Craver et Darden, 2013; Glennan, 1996; Machamer, Darden et Craver, 2000; pour une recension des écrits, voir Craver et Tabery, 2019). Je reprends ici les mots de Craver qui explique ce qu'il défend être un mécanisme :

Roughly, a mechanism is a collection of entities and activities organized such that they give rise to the behavior of a mechanism as a whole. Entities are objects, such as neurotransmitters and cells. These entities are characterized in terms of structural properties, such as their size, conformation, and material constituents, and in terms of the relations with other entities in the mechanism (their locations, relative motions, forces). Activities are the things the entities do, such as binding to receptors and generating action potentials. Activities, on my Woodwardian interpretation, are typically characterized by and a set of generalizations concerning the properties and organizational features required for an activity to occur and the consequences of such occurrence [...] The activities in neuroscience and physiology tend to be mechanistically explicable. (Craver, 2013, p. ??)

Dans le cadre de la MRC, c'est l'idée que les mécanismes sont « actifs », « localisés dans l'espace et dans le temps » et « organisés hiérarchiquement » qui vient spécifier les attributions fonctionnelles en neurosciences (Craver, 2001, p. 57). Pour mieux comprendre cette idée, il peut être utile de l'illustrer avec un exemple. Craver utilise le cas de la transmission synaptique rapide (SNC), dans laquelle le complexe des récepteurs à NMDA-ionophore est impliqué (voir la figure 2).

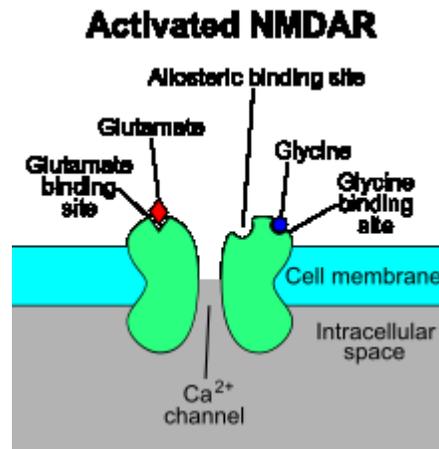


Figure 2. Récepteur NMDA

Source : NMDA Receptor, Wikipedia, consulté le 30 avril 2020.

La SNC joue un rôle important pour expliquer le système sensoriel et la mémoire. Le récepteur NMDA, quant à lui, joue un rôle de médiateur au sein de la SNC. L'activation du récepteur NMDA est l'une des (fonctions?) responsables de la transmission d'influx nerveux par communication intraneuronale. Cette communication se produit grâce à la transformation d'un signal chimique extracellulaire en un signal chimique intracellulaire. Selon la description neuroscientifique du phénomène, l'arrivée de l'influx nerveux au niveau de la terminaison nerveuse déclenche la libération de glutamate, un acide aminé excitateur (un neurotransmetteur). Le glutamate vient se fixer sur la membrane neurale et interagit avec le récepteur NMDA. Or le canal associé au NMDA est bloqué par un ion Mg^{2+} . Le canal ne peut être libéré que si une dépolarisation est produite à l'intérieur de la cellule grâce à l'activation des récepteurs AMPA situés près des canaux NMDA. Une fois les récepteurs AMPA activés et le Mg^{2+} retiré, l'entrée du canal NMDA se libère et les ions Na^+ et Ca^{2+} peuvent y entrer, et les ions K^+ en sortir. Ce flux d'ions, et plus particulièrement la circulation du Ca^{2+} , déclenche un potentiel d'action présynaptique (PPSE). La combinaison de plusieurs

PPSE provoque un potentiel d'action, responsable de la transmission de signaux nerveux (Bear *et al.*, 2007).

Avec cet exemple, on peut voir comment un mécanisme est actif spatialement et temporellement, en plus d'être hiérarchiquement organisé. Selon Craver, cette description inclut des entités comme une membrane cellulaire, un canal NMDA, les ions Ca^{2+} , etc. Cette description inclut également des activités comme la dépolarisation, l'entrée des ions dans le canal, etc. Ces entités et activités sont organisées de manière spatio-temporelle pour produire un phénomène, au sens où le phénomène est découpé en plusieurs étapes (d'abord la dépolarisation, puis l'entrée des ions Ca^{2+} , etc.). Ces entités et activités interagissent causalement pour produire un phénomène qui est hiérarchiquement supérieur : la communication intraneuronale (Craver, 2005). Dans un langage mécanique, cet exemple décrit donc le mécanisme du complexe des récepteurs à NMDA-ionophore. Le phénomène à expliquer est la transmission de Ca^{2+} à travers la membrane neuronale. On cherche donc à expliquer le mécanisme de la communication neuronale par transmission de Ca^{2+} (Craver, 2005, 2013). Ce phénomène intéresse les scientifiques parce qu'il contribue, entre autres, à une activité hiérarchiquement supérieure, la potentialisation à long terme (LTP), élément essentiel de la capacité de mémorisation spatiotemporelle. Du point de vue de Craver, un mécanisme est donc un ensemble d'entités et d'activités qui participent conjointement et causalement à la réalisation d'un comportement à expliquer.

Craver défend que cette notion de mécanisme trouvée dans les sciences biomédicales présuppose la notion de fonction (Craver, 2001; voir aussi Piccinini et Craver, 2011 pour le lien entre mécanisme et fonction). Lorsque l'on parle d'un processus ou d'une composante comme d'un mécanisme, c'est parce que l'on présuppose que ce processus ou cette composante participe à la réalisation de la capacité d'un système. Autrement dit, pour identifier un mécanisme au sein du système à l'étude, il faut avoir décidé quelle est la capacité du système que l'on veut étudier. Dans ce cadre, la fonction d'une

composante d'un système est l'activité réalisée par cette composante qui participe causalement au maintien de la capacité du système à l'étude. À cet effet, Craver écrit :

attributions of mechanistic role functions describe an item in terms of the properties or activities by virtue of which it contributes to the working of a containing mechanism, and in terms of the mechanistic organization by which it makes that contribution. (2001, p. 61).

Si l'on reprend l'exemple du complexe des récepteurs à NMDA-ionophore, on peut alors dire que la fonction de ce complexe est la transmission d'ions Ca^{2+} , et ce, en vue d'assurer la transmission neuronale. Le fait d'adopter le cadre des explications mécanistes pour rendre compte des attributions fonctionnelles permet d'ajouter une contrainte à ces attributions : elles doivent respecter les contraintes associées aux mécanismes. Ces contraintes impliquent que les éléments et activités d'un système doivent être localisés de manière spatio-temporelle, et qu'ils contribuent causalement à une capacité particulière du système à l'étude (Garson, 2016, p. 85).

Pour restreindre la trop grande libéralité de la RC, Craver insiste également sur le fait que les attributions fonctionnelles, en plus de devoir respecter les caractéristiques des mécanismes, peuvent être évaluées en fonction d'éléments de preuve. Il y aurait donc de bonnes attributions fonctionnelles et de mauvaises attributions fonctionnelles (Craver, 2001, p. 62, 2013, p. 141-143). Cette évaluation se fait au regard de la manière dont les mécanismes participent causalement à la réalisation de certaines activités à expliquer. Craver écrit :

I argue that functional attributions are contentful to the extent that they can be cashed out in a detailed description of how an item is organized into a higher-level mechanism, and that one has good evidence for one's functional description to the extent that one can show how the item is organized into the mechanism (Craver, 2013, p. 135).

L'évaluation empirique vise donc la manière dont les différentes composantes d'un système sont impliquées dans un mécanisme. Par exemple, une scientifique peut se demander quels sont les éléments de preuve qui permettent de confirmer que le récepteur NMDA participe causalement à la TPL? Si, après observation, l'action du récepteur NMDA ne participe pas à la diffusion de Ca^{2+} au sein des canaux de diffusion, alors l'attribution fonctionnelle « la fonction du récepteur NMDA est la communication neuronale » pourra être remise en question ou abandonnée.

Or, il est important de noter que même si l'approche de Craver vise à restreindre la libéralité de la RC par l'ajout de la notion de mécanisme, il n'y a pas à proprement parler d'éléments de preuve en faveur de la « bonne capacité » ou du « bon mécanisme » à expliquer. En ce sens, la question « Quelle est la « vraie » fonction d'un système? » ou « Quelle est la fonction « normale » d'un organe? » dépendra également de nos intérêts d'investigation, plutôt que d'une seule lecture des faits empiriques. Craver reste ainsi fidèle à la relativité des attributions fonctionnelles aux intérêts humains, caractéristique de la RC :

There is a fact of the matter about the causal connections composing an item's contextual causal nexus, but there is no uniquely correct fact of the matter as to which of those connections in the causal nexus ought to be highlighted in providing a response to a particular request for explanation. This is not a spooky variety of interest relativity, but the simple recognition that some aspects of the causal structure of the world are more interesting and relevant to our purposes than others. That should be explicit in our account of how we assign functions to the brain and its parts, it should not (be??) masked with obfuscating, vague and evidentially tenuous talk about contributions to fitness and propensities to survive or reproduce. (Craver, 2005, les italiques sont les miennes)

Pour Craver, la notion de fonction « normale » reflète les intérêts des scientifiques. Si un scientifique affirme que la fonction « normale » du complexe des récepteurs à

NMDA-ionophore est la transmission de Ca^{2+} , c'est-à-dire que le complexe est *censé* assurer la transmission de Ca^{2+} à travers la membrane neuronale, c'est parce que ce mécanisme l'intéresse (Craver, 2013). D'autres fonctions « normales » pourraient être attribuées à ce complexe si les intérêts des scientifiques divergeaient et que d'autres phénomènes étaient à expliquer. En ce sens, il est vrai de dire qu'un mécanisme peut expliquer la transmission de Ca^{2+} , mais qu'un autre mécanisme peut aussi expliquer son dérèglement. L'ajout du mécanisme à la RC ne répond pas entièrement à la question « quel mécanisme fonctionne ou dysfonctionne? ». Ce sont les intérêts des scientifiques qui permettent d'y répondre complètement. Or rappelons qu'en raison de cette relativité des attributions fonctionnelles aux intérêts humains, la RC faisait face au problème des traits dysfonctionnels et des traits non fonctionnels (voir la section 6.2.2). Est-ce à dire que la théorie de Craver, à l'image de la RC, ne peut pas répondre à ces problèmes?

Il est clair que pour Craver, seuls les intérêts humains peuvent départager la fonction de l'accident. Autrement dit, un « accident » pourrait très bien être perçu comme une fonction dans certains cadres explicatifs, si c'est ce qui intéresse les investigateurs. En cela, Craver ne répond pas vraiment à la critique des traits non fonctionnels, mais c'est parce qu'il ne perçoit pas cette relativité comme un problème. Cependant, Craver tente d'accommoder le problème des traits dysfonctionnels en appelant à la manière dont les mécanismes se comportent généralement dans ce type de système. Craver développe l'idée de généralisation pour rendre compte de la notion de dysfonction (pour une stratégie similaire, voir Hardcastle 1999, p. 36). Dans cette optique, un mécanisme dysfonctionnel serait un mécanisme qui s'écarte trop de ses activités typiques dans un système particulier (Craver, 2001, p. 72). Quant au jugement concernant ce qui est typique, le jugement de similarité entre les activités du mécanisme « dépend du niveau de tolérance convenu vis-à-vis la diversité des cas particuliers au sein des types » (Craver, 2001, p. 73). Ici, la typicité n'est pas naturelle, mais simplement le résultat d'une généralisation représentant la manière dont fonctionnent le plus souvent des

mécanismes. Identifier une dysfonction implique donc *un choix conventionnel* qui reflète la manière dont une communauté scientifique tolère la divergence vis-à-vis la typicité des mécanismes. C'est ce choix qui détermine si un mécanisme est fonctionnel ou dysfonctionnel. Selon Craver, on peut parler de dysfonction si certains mécanismes s'écartent trop de la manière dont ils participent *généralement* à la réalisation d'une activité dans un système. Par exemple, un cœur pourra être dit « dysfonctionnel » si la manière dont il pompe le sang diverge trop de la manière dont les cœurs humains pompent généralement le sang pour assurer la circulation sanguine. Cet écart trop grand par rapport au fonctionnement généralisé du cœur est jugé selon les conventions d'une communauté scientifique, en fonction de ses intérêts. Dans cette optique, l'idée de généralisation vient contraindre la trop grande libéralité de la RC et répondre au problème des traits dysfonctionnels, sans pour autant rejeter l'idée que les attributions fonctionnelles dépendent des intérêts humains.

Parce que les attributions fonctionnelles et l'identification de dysfonctions dépendent des intérêts humains, Craver adopte ce qu'il appelle un « perspectivalisme » (Craver, 2013; voir aussi Craver, 2001, p. 71). Il écrit contre l'approche ES : « The perspectivalist refuses to pass the buck for these normative judgments to evolutionary biology. When we make such normative judgments, says the perspectivalist, we are responsible for justifying them. Evolution cannot do that work for us. » (Craver, 2013, p. 149). Pour Craver, la normativité des fonctions provient des intérêts humains, et non pas de la nature comme le souhaitaient les approches OB ou ES défendues respectivement par Boorse et Wakefield. Or comme déjà suggéré, la dépendance des attributions fonctionnelles aux intérêts humains peut entraîner certains problèmes, même si la notion de mécanisme et celle de généralisation ont été ajoutées à l'analyse de la RC. D'abord, de l'aveu même de Craver, le perspectivalisme entraîne la possibilité que des biais problématiques façonnent la manière dont nous traçons la ligne entre le fonctionnel et le dysfonctionnel (Craver, 2001, p. 73). Une autre implication, toujours notée par Craver, est que si ce sont les pathologies qui intéressent les

scientifiques, des mécanismes qui expliquent les pathologies peuvent être découverts. Dans le cadre de la MRC, les composantes et processus qui expliquent les pathologies peuvent être qualifiés de fonctionnels, puisqu'ils participent à maintenir les pathologies à l'étude (Craver, 2001, p. 67). Puisque la théorie de Craver ne repose pas sur des bases qui permettent d'identifier *la* fonction normale d'un système indépendamment des intérêts humains, il s'en suit ce résultat contre-intuitif⁶.

Dans tous les cas, ce qu'il faut retenir de l'approche de Craver, c'est que même si la RC a été considérablement modifiée et améliorée pour restreindre sa trop grande libéralité par les notions de mécanismes et de généralisation, la normativité ne provient pas de la nature. La normativité provient des intérêts humains. Les fonctions sont étudiées une fois que ces états ont été identifiés comme ayant de la valeur pour l'investigation. Ainsi, dans la MRC, outre la caractérisation d'un mécanisme offerte par Craver et l'idée de généralisation, il n'y a pas de contraintes autres que les intérêts des chercheurs pour venir limiter les attributions fonctionnelles, et plus spécifiquement pour démarquer la fonction de la dysfonction. Cette relativité est perçue par plusieurs comme problématique (voir Von Eckardt et Poland, 2004, p. 582; Garson, 2013, 2019).

⁶ Une possible solution est à trouver chez Nervi (2010). Pour Nervi, lorsque les pathologistes étudient les mécanismes qui expliquent les pathologies, ils le font avec en arrière-plan *le présupposé de ce qu'est la fonction normale*. Même si l'objet d'étude principal des pathologistes est le mécanisme de la pathologie, le cadre général explicatif contiendrait aussi l'idée de fonction normale. Ainsi, le mécanisme de la pathologie ne pourrait pas être dit fonctionnel, puisque le trait fonctionnel est déjà présupposé dans ce cadre d'analyse. Cela irait à l'encontre de ce qu'affirme Craver, qui accepte que des pathologies puissent être qualifiées de fonctionnelles. Or, cette solution serait peut-être plus intuitive et plus conforme à la manière dont les pathologistes parlent de mécanismes et de fonctions (les pathologistes semblent en effet parler de mécanismes, mais pas vraiment de mécanismes fonctionnels ou de fonctions lorsqu'ils décrivent les composantes de la pathologie). Ainsi, même si théoriquement, cet aspect du problème des traits dysfonctionnels serait réel, dans la pratique, il ne se présenterait pas vraiment en raison de la manière dont le système à l'étude est découpé par les scientifiques.

6.3 La solution de Roe et Murphy : attributions fonctionnelles et idéalizations

Kelly Roe et Dominic Murphy (2011; voir aussi Murphy, 2010, 2021) ont récemment présenté une défense de la RC en psychiatrie, adoptant l'élaboration mécaniste de la MRC proposée par Craver. Cette défense s'accompagne d'un amendement à la MRC : spécifier la notion de généralisation par la notion d'idéalisation (voir aussi Schaffner, 1986, 1993, pour une idée similaire selon laquelle il y aurait en médecine des théories se présentant comme des modèles; Lemoine, 2014 pour une présentation et une discussion). Leur allégeance à la MCR (qu'ils appellent l'approche de la capacité systémique) s'explique par le fait que cette théorie décrirait mieux comment la médecine, et par là la psychiatrie, utilise le langage fonctionnel. Selon Roe et Murphy, de manière générale, la médecine s'intéresse peu au caractère adaptatif (au sens évolutionniste) des traits et des comportements qu'elle étudie (à l'exception bien sûr de la médecine et de la psychiatrie évolutionnaire, voir chapitre 3). Roe et Murphy décrivent les attributions fonctionnelles en médecine, et plus précisément la manière dont une dysfonction est identifiée :

How is dysfunction determined? By the use of a biomedical concept of normality that is an *idealized description of a component of a biological system in an unperturbed state that may never be attained in actual systems*. It does not rest on the failure of a biological part to function as its ancestors did, but *by its failure to be close enough to the causal contribution of the analogous part in the idealized overall system*. (Roe et Murphy, 2011, p. 222, les italiques sont les miennes)

Dans cette optique, des fonctions peuvent être identifiées comme des contributions causales des sous-systèmes à des capacités du système général, à l'image de ce que prescrit la MRC. Or, l'originalité de Roe et Murphy réside dans leur tentative de rendre

compte de la notion de dysfonction dans ce cadre. Pour les deux auteurs, les fonctions *normales* sont idéalisées : elles représentent ce que serait un mécanisme qui participe causalement à maintenir une activité particulière de manière efficace, sans perturbation (voir aussi Ereshefsky, 2009, p. 223; Wachbroit, 1994, p. 237 pour le lien entre normalité et idéalisation). Une dysfonction est alors un état qui s'écarte trop du modèle de fonctionnement idéalisé.

Pour clarifier davantage la proposition de Murphy, il peut être opportun de développer brièvement la notion d'idéalisation. En science, une idéalisation peut être définie de manière très générale comme une distorsion intentionnelle d'un phénomène à l'étude (p. ex. : les agents rationnels en économie, un plan sans friction en physique, etc.) (voir Potochnik, 2017, p. 2; Weisberg, 2007). Cette distorsion est généralement explicable par l'un des buts de la science, celui de la représentation des phénomènes à l'étude par des théories (Weisberg, 2007). Deux types d'idéalisation peuvent être distingués : l'idéalisation aristotélicienne et l'idéalisation galiléenne. L'idéalisation aristotélicienne est une abstraction où l'on simplifie le phénomène en enlevant des éléments qui ne sont pas considérés comme pertinents. Cette méthode permet d'isoler des propriétés qui sont nécessaires à la résolution du problème scientifique à l'étude. Un exemple classique d'idéalisation aristotélicienne est la loi de Boyle-Mariotte, qui décrit comment la pression d'un gaz a tendance à augmenter lorsque le volume du récipient diminue (et inversement, lorsque le volume augmente, la pression du gaz diminue). La loi s'applique aux gaz idéaux, de masse et de température fixes. Pour parvenir à cette loi, les physiciens ont assumé que les molécules de gaz n'entraient pas en collision les unes avec les autres. Or dans la réalité, les molécules entrent en collision, mais que cette collision n'a pas d'impact sur la loi. Les collisions sont donc considérées comme non pertinentes à la résolution du problème et exclues de la représentation. Quant à l'idéalisation galiléenne, elle est une représentation volontairement faussée d'un phénomène, au sens où des paramètres du phénomène ont été simplifiés pour le rendre calculable (McMullin, 1985). L'idéalisation galiléenne est employée dans un but de

« dé-idéalisation » : les scientifiques peuvent réintroduire les paramètres modifiés pour une représentation plus complète du phénomène au fur et à mesure que le phénomène devient calculable. Par exemple, en mécanique quantique, les chimistes-informaticiens ont d'abord calculé les fonctions d'onde moléculaires (Ψ) de manière approximative (la fonction d'onde moléculaire est la représentation de l'état quantique d'un système). C'est avec le développement des ordinateurs électroniques que les chercheurs ont pu calculer de manière plus précise les fonctions d'onde moléculaires. Or ce calcul est toujours approximatif, parce qu'à partir des équations disponibles, seule la fonction d'onde de l'atome d'hydrogène et de l'ion moléculaire H_2^+ peuvent être calculées avec précision. Le reste des calculs se fait par approximation et produit une solution rapprochée, mais pas exacte (voir Weisberg, 2007 pour l'exemple et une discussion).

Dans les deux cas, il s'en suit qu'une idéalisation est une représentation qui ne vise pas nécessairement la vérité, mais d'autres buts épistémiques (comme la compréhension, la prédiction, etc.) (p. ex. : Frigg, 2012). Si l'on revient au cas de la médecine et de la psychiatrie, ces types de procédés semblent s'appliquer. Prenons l'exemple du cœur. Les manuels de physiologie postulent que la fonction du cœur est de pomper le sang, selon une fraction d'éjection entre 50 et 70 %. C'est de cette manière que la circulation sanguine peut être maintenue de manière efficace pour l'être humain. La fonction normale du cœur est ainsi *idéalisée* selon la manière dont, causalement, le cœur permettrait une circulation sanguine adéquate. Dans une population humaine, il n'est donc pas requis que l'ensemble de cette population réalise cette fonction, ni même une majorité, pour que l'on puisse dire qu'il s'agit de la fonction du cœur. L'idéalisation est une représentation du fonctionnement du cœur dans une situation idéale sans perturbation qui peut servir les buts de la médecine (p. ex. : poser un diagnostic au bon moment, éviter le développement de symptômes préjudiciables, développer et prescrire traitements pour guérir les patients) (voir aussi Murphy, 2010, cette idée permet de rendre compte du problème des « pathologies universelles », comme les caries et les irritations pulmonaires, auquel faisait face Boorse, voir le chapitre 1).

L'idée que le fonctionnement normal d'un organisme se traduise par des idéalizations est développée dans un autre article de Murphy (2011), dans lequel il soutient que la santé mentale serait une forme d'idéalisation (et le trouble mental une déviation vis-à-vis cet état idéalisé). Selon Murphy, les psychiatres et les neuropsychologues utilisent implicitement une théorie des fonctions mentales normales basée sur des structures cognitives qui seraient un ensemble d'idéalizations. En ce sens, l'ensemble des fonctions qui seraient associées à la cognition, à la perception, à la régulation de l'humeur (etc.) serait idéalisé. Qui plus est, pour Murphy, cette idéalisation n'est pas de nature statistique, comme Boorse le soutenait (voir le chapitre 1). Murphy écrit :

[...] when we say that an organ is normal, we employ a biomedical concept of normality that is neither normative nor statistical. Rather, *it is an idealized description of a component of a biological system in an unperturbed state that may never be attained in actual systems*. But it is the account of the organ that gets into the physiology textbook [...] It draws all its authority from its predictive and explanatory utility : against the background of assuming normal heart function, for example, we account for variation in actual hearts (a particular rhythm, say), by citing the textbook rhythmic pattern (which may be very unusual statistically) and identifying other patterns as arrhythmic. Likewise, psychiatrists and neuropsychologists assume a theory of normal function of cognitive structures that is a set of idealizations. (Murphy, 2010, p. 447, les italiques sont les miennes)

La proposition de Roe et Murphy demeure très « schématique » et demanderait à être élaborée davantage dans de futurs travaux. En effet, Murphy s'intéresse plutôt aux idéalizations utilisées pour expliquer les mécanismes impliqués dans des troubles mentaux particuliers, comme la dépression ou la schizophrénie, plutôt que pour expliquer les concepts de santé mentale et de trouble mental en général. Qui plus est, comme le note Wachbroit (1994) au sujet de la théorie sur laquelle la proposition de Murphy est basée, en physique par exemple, où les idéalizations sont nombreuses, on ne parle jamais de « système planétaire normal » ou d'« interactions magnétiques normales », alors que le concept de normalité est central pour la santé et la maladie.

Plus de travail doit donc être fait par rapport à la manière dont des éléments évaluatifs relevant de la normalité peuvent être intégrés à une idéalisation⁷. Néanmoins, l'idée d'idéalisation défendue par Roe et Murphy en complément à la MRC de Craver représente une piste intéressante pour rendre compte de l'identification de dysfonctions en médecine et en psychiatrie, et ce, toujours dans le but de dépasser le problème des traits dysfonctionnels.

Même si la notion d'idéalisation est intéressante, un point sur lequel il convient d'insister à nouveau est la relativité de ce type d'attribution fonctionnelle aux intérêts humains. Même si la proposition de Roe et Murphy est une amélioration par rapport à l'apport original de Cummins, parce qu'elle intègre à la fois la notion de mécanisme et celle d'idéalisation (par contraste avec la notion moins spécifique de généralisation), elle ne constitue pas une solution entière au problème de la trop grande libéralité de la CR. En effet, nous nous retrouvons tout de même face à des attributions fonctionnelles qui dépendent des intérêts des scientifiques pour certaines capacités qui précèdent l'attribution des fonctions aux traits et aux comportements qui contribuent causalement à la capacité à expliquer. Roe et Murphy sont conscients de cette implication. Ils défendent que, puisque les valeurs sont toujours présentes dans les attributions fonctionnelles en médecine, « science does not uncover dysfunction in a way that is independent of our value judgments; science is directed by those value judgments » (Roe et Murphy, 2011). On demeure donc dans ce que Craver a appelé le perspectivalisme.

⁷ Plus de travail doit également être mené pour justifier le lien entre les neurosciences cognitives et la psychiatrie. En effet, Murphy (2006) plaide pour que la psychiatrie base ses théories et idéalizations sur les neurosciences cognitives, parce que capables d'offrir des explications mécanistes et causales des troubles mentaux. Or cette thèse ne fait pas l'unanimité et demande à être davantage justifiée. Quoiqu'il en soit, de manière générale, la proposition de Murphy est intéressante et constitue une base pour pousser plus avant le lien entre santé mentale et idéalisation.

Autrement dit, l'approche de Roe et Murphy conserve la dépendance de l'identification des traits fonctionnels et dysfonctionnels aux intérêts humains. En effet, même si l'on accepte que les fonctions normales soient des idéalizations qui représentent le fonctionnement d'organes et de systèmes dans des états idéaux non perturbés, la nature de ces états et la manière dont ceux-ci seront idéalisés demandent elles-mêmes à être déterminées par rapport à des intérêts particuliers. Par exemple, la circulation sanguine est une capacité qui intéresse les scientifiques parce qu'elle permet de maintenir les êtres humains en vie, mais aussi d'assurer que les êtres humains puissent maintenir certaines activités valorisées (p. ex. : faire du sport). Le modèle idéalisé de la circulation sanguine n'est donc pas indépendant des intérêts humains. La normativité provient donc toujours du fait que les scientifiques sont intéressés par certaines capacités qui permettent aux êtres humains de réaliser des activités valorisées⁸.

Comme je l'ai déjà souligné dans la section précédente (6.3), l'une des implications problématiques de cette relativité est la possible présence de biais et de valeurs préjudiciables dans les attributions fonctionnelles, comme l'a noté Craver. Dans le cadre de la MRC, et même si la notion d'idéalisation est ajoutée, il n'y a rien qui empêche de simples variations au sein de la population humaine d'être considérées comme des dysfonctions. Par exemple, l'homosexualité pourrait être considérée comme une dysfonction si les valeurs des scientifiques sont influencées par des présupposés hétéronormatifs. Si ce qui intéresse les scientifiques est l'hétérosexualité

⁸ Ici, un parallèle peut être tracé avec la manière dont le handicap est conceptualisé dans les manuels officiels de classification des handicaps (comme par exemple le *International Classification of Functioning, Disability and Health* [ICF]). Dans ce manuel, un handicap est compris selon le modèle biopsychosocial, selon lequel le handicap est causé par l'interaction d'un ensemble de facteurs physiologiques et sociaux. Au sein de ce modèle, le handicap peut être causé, du moins en partie, par la perte d'une capacité nécessaire à l'exécution d'une activité valorisée. Par contre, ce parallèle a ses limites : l'ICF distingue la pathologie ou le trouble, qui seraient purement physiologiques, du handicap. Ce que je défends dans cette thèse est que le trouble (mental) et la dysfonction sont eux aussi étudiés en fonction des intérêts humains et des conditions valorisées.

parce que c'est l'orientation sexuelle valorisée dans une société donnée ou même dans une communauté scientifique donnée, les mécanismes qui sont responsables de l'orientation sexuelle pourraient être étudiés et idéalisés. L'homosexualité pourrait être perçue comme un comportement dysfonctionnel, parce que divergeant trop des mécanismes « typiques » de l'orientation sexuelle souhaitée. Autrement dit, il serait possible d'idéaliser les mécanismes qui expliquent l'hétérosexualité et de concevoir les mécanismes qui expliquent l'homosexualité comme des dysfonctions qui dévient trop de l'idéalisation. Cette implication paraît problématique. En effet, si les intérêts humains sont problématiques, alors les attributions fonctionnelles seront guidées par ses valeurs problématiques. Sans moyens pour réguler cette influence, le perspectivalisme est insatisfaisant, parce qu'il ne permet pas de contraindre la manière dont les intérêts humains viennent guider les attributions fonctionnelles, ouvrant la porte à l'influence de valeurs problématiques comme l'homophobie. Ainsi, dans leur état actuel, les théories de la fonction MRC, qu'elles s'accompagnent de la notion de généralisation ou d'idéalisation, n'offrent donc pas de solutions ou d'outils pour répondre à cette implication problématique de la relativité des attributions fonctionnelles aux intérêts humains.

Malgré cette limite, il est également important de reconnaître qu'adopter une théorie perspectivaliste de la fonction comme la MRC présente au moins un avantage majeur. Cela permet de rendre compte de l'influence des valeurs sur les attributions fonctionnelles en science. Comme mentionné précédemment, plusieurs auteurs ont en effet défendu que la MRC décrivait mieux la manière dont les scientifiques attribuaient des fonctions dans certains domaines de recherche comme la physiologie et les neurosciences. Qui plus est, et comme je l'ai défendu dans les chapitres 2 et 3, ils semblent que même les théories de la fonction qui se présentent comme indépendantes des valeurs impliquent au moins certains choix normatifs pour que des attributions fonctionnelles soient possibles. Si ces auteurs ont raison, c'est donc que les valeurs

influencent les attributions fonctionnelles dans certains domaines de recherche. Le perspectivalisme fournit le cadre pour penser cette influence.

Dans la section suivante, je vais défendre que cette limite ne doit pas nous conduire au rejet de la MRC, mais plutôt à en proposer un complément : l'objectivité sociale. Je vais proposer que les valeurs qui guident les attributions fonctionnelles de la MRC doivent être régulées par les normes de l'objectivité sociale pour que cette théorie soit pleinement satisfaisante. Ainsi, bien que dans les théories de la fonction MRC, la normativité provienne du social et non pas de la nature (comme le souhaitait l'approche OFN), cette normativité peut tout de même être objective et satisfaisante pour une définition du concept de trouble mental. Les attributions fonctionnelles, même si dépendantes des intérêts humains, pourraient donc être objectives.

6.4 Les attributions fonctionnelles et les valeurs en science

Comme je l'ai mis en évidence dans les sections précédentes, il est généralement défendu que le problème principal de la RC et de sa version améliorée, la MRC, est sa trop grande libéralité. Plus spécifiquement, c'est la relativité des attributions fonctionnelles aux intérêts humains qui est perçue par plusieurs comme une limite importante. Dans cette section, je vais défendre que la relativité des attributions fonctionnelles aux intérêts humains n'est pas un problème en soi. Le principal problème de la MRC réside plutôt dans le fait que cette théorie n'offre pas d'outils pour que les intérêts, valeurs, et possibles biais problématiques qui influencent les attributions fonctionnelles soient identifiés et régulés. Cette thèse se base sur l'argument contre l'idéal de la science indépendante des valeurs présenté au chapitre 4 et sur la défense de son alternative, l'objectivité sociale. Rejeter l'idéal de la science indépendante des valeurs au profit de l'idéal de l'objectivité sociale permettrait une

meilleure garantie que les valeurs présentes dans les attributions fonctionnelles soient explicitées et contrôlées, plutôt qu'occultées. Dans ce qui suit, je vais brièvement rappeler cet argument et montrer comment il s'applique également aux théories de la fonction MRC.

Dans le chapitre 4, j'ai montré que l'OFN cherchait à élaborer une définition du trouble mental avec une notion de dysfonction indépendante des valeurs parce que cette approche s'inscrivait dans ce qu'on appelait l'idéal de la science indépendante des valeurs. Dans cette optique, la notion de dysfonction devait être indépendante des valeurs (ou des intérêts humains), parce que seuls les concepts scientifiques qui ne font pas intervenir les valeurs sociales et culturelles pouvaient être dits « objectifs », et donc « scientifiques ». J'ai remis en question cet idéal, suivant deux formes de critiques : la critique sémantique et la critique méthodologique. Or comme je l'ai montré, rejeter cet idéal implique la présence probable de valeurs non épistémiques dans nos concepts et théories scientifiques. Certaines de ces valeurs peuvent être problématiques. Il faut donc trouver un moyen de pouvoir identifier et contrôler les valeurs pour être en mesure de freiner l'influence de valeurs problématiques dans la production de connaissances scientifiques. J'ai soutenu que les normes de l'objectivité sociale se présentaient comme un bon moyen d'identifier et de critiquer les valeurs. L'OS est donc une bonne alternative pour penser un concept de trouble mental socialement construit et objectif.

Je crois que la même réflexion peut être appliquée aux théories de la fonction normale et aux critiques qui ont visé la trop grande libéralité de la RC et de la MRC. Selon moi, la RC et la MRC n'ont pas été retenues pour fonder une définition du trouble mental en raison de la présence suspecte des valeurs et des intérêts humains dans les attributions fonctionnelles. Comme je l'ai montré avec l'exemple de l'homosexualité, il y a de bonnes raisons de demeurer suspects vis-à-vis ces valeurs. Or si les critiques de l'idéal de la science indépendante des valeurs sont justes, il n'y a pas de bonnes raisons de s'opposer a priori à la RC et à la MRC parce qu'elles sont perspectivalistes.

Le fait que des valeurs non épistémiques influencent les attributions fonctionnelles n'est pas un problème en soi, tant et aussi longtemps que les normes de l'objectivité sociale sont respectées. Autrement dit, si des structures sont mises en place pour que soient identifiées et critiquées les valeurs, il y a beaucoup moins de chance que leur influence soit problématique. Je soutiens donc que la principale limite de la MRC n'est pas la relativité des attributions fonctionnelles aux intérêts humains, mais le fait qu'il n'y ait pas d'outils pour contraindre et réguler cette relativité. L'OS peut fournir cette contrainte.

Généralement, les critiques de l'approche OFN plaident également pour l'abandon d'une théorie de la fonction pour fonder le concept de trouble mental. Or, je soutiens que rejeter l'OFN et l'idéal de la science indépendante des valeurs n'implique pas nécessairement de se priver d'une théorie de la fonction normale. Cette théorie peut être perspectivaliste, si elle s'accompagne des normes de l'OS. Dans le chapitre suivant, je vais m'intéresser à la manière dont les normes de l'objectivité sociale pourraient s'appliquer en psychiatrie, mais d'abord, dans la section suivante, je vais clarifier comment la théorie de la fonction MRC peut constituer une solution au desideratum d'objectivité *empirique* du concept de trouble mental.

6.5 Objectivité empirique et mécanismes causaux

Dans les sections précédentes, j'ai montré comment il était possible de contraindre la liberté des attributions fonctionnelles décrites par la RC par l'ajout des notions de mécanisme, et j'ai suggéré que la MRC serait moins sujette à l'influence de valeurs problématiques et par là plus satisfaisante si les normes de l'objectivité sociale étaient

appliquées. Pour compléter ce chapitre, je soutiendrai qu'adopter la théorie de la fonction MRC pourrait apporter le contenu empirique dont le trouble mental a besoin pour être dit *empiriquement* objectif.

Comme je l'ai exposé dans la section 6.3, l'ajout de la notion de mécanisme à la théorie de la fonction RC permet de contraindre la libéralité de la MRC. Or elle permet aussi de donner un contenu empirique plus solide à cette théorie. En effet, rappelons que selon Craver, une attribution fonctionnelle peut être dite vraie ou fausse dans la mesure où une composante d'un système, par ses activités ou ses propriétés, participe causalement à la réalisation d'une capacité hiérarchiquement supérieure à expliquer. Cette contribution causale d'une composante, c'est-à-dire sa fonction, peut être vérifiée empiriquement. Ainsi, si une composante ne participe pas au maintien de la capacité d'un système, alors il n'est pas empiriquement correct de lui attribuer une fonction au sein de ce système. En ce sens, une attribution fonctionnelle peut être dite empirique, parce qu'elle est testable empiriquement. Dans le cadre perspectivaliste de la MRC, l'attribution fonctionnelle n'est donc pas le reflet pur et simple des intérêts humains, puisqu'une fois ces intérêts circonscrits, des éléments de preuve viennent appuyer ou informer ces attributions.

Par contre, une clarification s'impose. Selon la MRC, ce qui n'est pas empirique dans ces attributions fonctionnelles est le choix de la capacité à l'étude. Le choix de s'intéresser à une capacité particulière reflète bel et bien les intérêts humains. Par exemple, si des scientifiques s'intéressent à certaines capacités particulières de l'esprit humain comme la perception ou la régulation des émotions, cet intérêt n'est pas motivé par des observations empiriques concernant l'esprit humain. C'est plutôt que ces capacités intéressent les chercheurs. Or une fois ces capacités déterminées, il devient possible d'identifier les entités et activités qui participent causalement au maintien de ces capacités. C'est donc qu'une fois les capacités identifiées, des informations empiriques par rapport aux fonctions des entités et activités impliquées causalement

dans la production et le maintien de ces capacités peuvent être amassées. Prenons par exemple le cas de la régulation des émotions. Cette capacité peut être d'intérêt pour les chercheurs, parce qu'elle permet aux êtres humains de conserver une humeur stable dans leur vie quotidienne. Cette capacité est donc valorisée parce qu'elle peut contribuer à maintenir ou améliorer la qualité de vie des êtres humains (voir Bandura *et al.*, 2003). Une fois cette capacité choisie, il est possible d'observer les mécanismes de la régulation des émotions et de décomposer ces mécanismes en diverses composantes et activités spatiotemporellement localisées. Il sera alors possible d'attribuer des fonctions à ces composantes. Bien que les mécanismes neuronaux impliqués dans la régulation des émotions ne soient pas parfaitement connus, il est possible d'esquisser la manière dont les attributions fonctionnelles pourraient prendre forme une fois la capacité « régulation émotionnelle » identifiée comme importante. L'amygdale, structure située dans la partie préfrontale du lobe temporal semble jouer un rôle primordial dans la régulation des émotions, principalement de la peur et de l'agressivité. Sans entrer dans les détails de ces hypothèses, il est possible de dire que l'une des fonctions de l'amygdale est la régulation de la peur et de l'agressivité (Bear *et al.*, 2007). Cette hypothèse peut être testée empiriquement, par exemple en utilisant des modèles animaux dont l'amygdale présenterait une lésion, ou par une stimulation intracrânienne de l'amygdale chez des cobayes animaux. Cette hypothèse peut également être testée par l'observation de patients humains dont l'amygdale présente une lésion ou par IRMf. Si, à la suite de ces observations, il s'avère que l'amygdale ne participe pas causalement au traitement de la peur et de l'agressivité, alors l'hypothèse « la fonction de l'amygdale est la régulation des émotions » devra être infirmée. Or ce qui ne peut pas être testé empiriquement, c'est l'affirmation « la capacité de régulation de la peur et de l'agressivité est normale chez l'être humain » (au sens où la normativité serait indépendante des intérêts humains). Rien dans la MRC ne permet de postuler que la normalité peut être découverte empiriquement. Les capacités à l'étude sont déterminées en fonction des intérêts humains. Les fonctions « normales » sont donc elles aussi le reflet de ces intérêts. Ce n'est qu'une fois que des capacités ont été établies

comme importantes que les attributions fonctionnelles peuvent être dites empiriques, parce que testables empiriquement en fonction de la manière dont les composantes d'un système participent à la réalisation d'une capacité. Ainsi, bien que ce ne soit pas l'entièreté de la MRC qui relève de l'empirique, le fait que les attributions fonctionnelles puissent être testées empiriquement une fois les capacités de l'esprit humain choisies me semble suffisant pour affirmer que du contenu empirique peut être associé à ces fonctions. Dire que le trouble mental est une dysfonction pourrait donc être objectivement empirique si les valeurs en jeu sont soumises aux normes de l'objectivité sociale.

Conclusion : fonctions comme rôle causal mécanique et perspectivalisme

Comme je l'ai expliqué, une théorie de la fonction perspectivaliste peut être à la fois objective socialement *et* empiriquement. Comme précisé dans la section précédente, la théorie de la fonction MRC permet de donner un contenu empirique à un concept de trouble mental défini comme une dysfonction (principalement grâce à la notion de mécanisme, testable empiriquement). Or cette dysfonction n'est pas à interpréter comme une notion indépendante des valeurs et intérêts humains. Le perspectivalisme, caractéristique de la MRC, implique une relativité des attributions fonctionnelles aux intérêts humains. Il n'en demeure pas moins qu'une fois que certaines capacités de l'esprit humain sont identifiées comme les « bonnes » capacités, ou celles qu'il est important de pouvoir maintenir (parce que, par exemple, elle contribue au bien-être des individus, à leur capacité d'action, à minimiser leur souffrance, etc.), alors ces attributions fonctionnelles sont testables empiriquement.

Cependant, une implication problématique de ce perspectivalisme est qu'il devient possible que des conditions dévalorisées (p. ex. : l'homosexualité) soient conçues comme des troubles mentaux. Comme le relève Craver, puisque les attributions fonctionnelles relèvent en partie des valeurs et des intérêts humains, alors des biais problématiques peuvent s'y glisser. Si, comme je l'ai illustré, ces valeurs sont hétéronormatives ou homophobes, alors il est possible que l'homosexualité soit considérée comme une dysfonction, et donc comme un trouble mental. Or, prétendre que les fonctions « normales » reflètent la normativité de la nature n'est pas une option pour barrer la route à ce type de présupposés problématiques (comme on l'a vu, les approches de la fonction indépendante des valeurs échouent). Je pense donc que l'approche de l'objectivité sociale est le meilleur espoir pour faire face à de potentiels biais problématiques venant influencer les attributions fonctionnelles en psychiatrie, puisqu'elle permet de rendre explicites ces biais et de les critiquer. L'objectivité sociale liée à la théorie de la fonction MRC permet donc d'imaginer un concept de trouble mental socialement et empiriquement objectif. Dans le chapitre suivant, j'esquisserai la manière dont l'objectivité sociale pourrait prendre place dans les institutions de psychiatrie nord-américaines.

Références du chapitre 6

- Bandura, A., Caprara, G. V., Barbaranelli, C., Gerbino, M. et Pastorelli, C. (2003). Role of affective self-regulatory efficacy in diverse spheres of psychosocial functioning. *Child Development*, 74(3), 769-782.
- Bear, M. F., Connors, B. W. et Paradiso, M. E. (2007). *Neurosciences : À la découverte du cerveau*. (3^e édition). Rueil-Malmaison : Éditions Pradel.
- Bechtel, W. et Abrahamsen, A. (2005). Explanation : A mechanist alternative. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421-441.
- Bechtel, W. et R. C. Richardson (1993/2010). *Discovering complexity : Decomposition and localization as strategies in scientific research*. (2^e édition). Cambridge, MA : MIT Press/Bradford Books.
- Campaner, R. (2011). *Understanding mechanisms in the health sciences*. *Theoretical Medicine and Bioethics*, 32(1), 5-17.
- Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, 68(1), 53-74.
- Craver, C. (2005). Functions and mechanisms in contemporary neuroscience. Dans P. Poirier, L. Faucher, E. Racine et E. Ennan (dir.), *Des neurones à la conscience : Neurophilosophie et philosophie des neurosciences*. Bruxelles : De Boeck Université.
- Craver, C. F. (2007). *Explaining the brain : Mechanisms and the mosaic unity of neuroscience*. New York : Oxford University Press
- Craver, C. F. (2013). Functions and mechanisms : A perspectivalist view. Dans P. Huneman (dir.), *Functions : Selection and mechanisms* (p. 133-158). Dordrecht : Springer.

- Craver, C. F. et L. Darden (2013). *In search of mechanisms : Discoveries across the life sciences*. Chicago : University of Chicago Press.
- Craver, C. et Tabery, J. (2019). Mechanisms in Science. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>
- Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA : Bradford/ MIT Press.
- Davies, P. S. (2001). *Norms of nature : Naturalism and the nature of functions*. Cambridge, MA : MIT Press.
- Eckardt, B. V. et Poland, J. S. (2004). Mechanism and explanation in cognitive neuroscience. *Philosophy of Science*, 71(5), 972-984.
- Garson, J. R. (2006). *Psychiatric disorders and biological dysfunctions : Some philosophical questions concerning psychiatry* (thèse de doctorat). University of Texas at Austin.
- Garson, J. (2013). The functional sense of mechanism. *Philosophy of Science*, 80(3), 317-333.
- Garson, J. (2016). *A critical overview of biological functions*. Cham : Springer International Publishing.
- Garson, J. (2019). There are no ahistorical theories of function. *Philosophy of Science*, 86(5), 1146-1156.
- Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 49-71.
- Hardcastle, V. G. (1999). Understanding functions. Dans V. G. Hardcastle (dir.), *Where biology meets psychology* (p. 27-43). Cambridge, MA : MIT Press.

- Hardcastle, V. G. (2002). On the normativity of functions. Dans A. R. Ariew, R. Cummins et M. Perlman (dir.). *Functions : New essays in the philosophy of psychology and biology* (p. 144-156). Oxford : Oxford University Press.
- Kitcher, P. (1993). Function and design. *Midwest Studies in Philosophy*, 18, 379-397.
- Lange, M. (2007). The end of disease. *Philosophical Topics*, 35, 265-292.
- Lemoine, M. (2014). Trois conceptions sémantiques des théories en médecine. *Lato Sensu, Revue de la Société de philosophie des sciences*, 1(1), 1-11.
- MacDougall, D. R. (2020). The ends of medicine and the experience of patients. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 45(2), 129-144.
- Machamer, P., Darden, L. et Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1-25.
- McCaffrey, J. B. (2015). The brain's heterogeneous functional landscape. *Philosophy of Science*, 82(5), 1010-1022.
- McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A*, 16(3), 247-273.
- Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, 56, 288-302.
- Murphy, D. (2010). Complex mental disorders : Representation, stability and explanation. *European Journal of Analytic Philosophy*, 6(1), 28-42.
- Murphy, D. (2011). Conceptual foundations of biological psychiatry. Dans F. Gifford (dir.), *Philosophy of Medicine* (p. 425-451). North Holland : Elsevier.
- Murphy, D. (2021). Function and dysfunction. Dans L. Faucher et D. Forest (dir.), *Defining mental disorders : Jerome Wakefield and his critics* (p. 251-266). Cambridge, MA : MIT Press.

- Piccinini, G. et Craver, C. F. (2011). Integrating psychology and neuroscience : Functional analyses as mechanism sketches. *Synthese*, 183(3), 283-311.
- Roe, K. et Murphy, D. (2011). Function, dysfunction, adaptation? Dans P. Adriaens et A. Block (dir.), *Maladapting minds* (p. 216-237). New York : Oxford University Press.
- Schaffner, K. F. (1986). Exemplar reasoning about biological models and diseases : A relation between the philosophy of medicine and philosophy of science. *The Journal of medicine and philosophy*, 11(1), 63-80.
- Schaffner, K. F. (1993). *Discovery and explanation in biology and medicine*. Chicago : University of Chicago Press.
- Wachbroit, R. (1994). Normality as a biological concept. *Philosophy of Science*, 61, 579-591.
- Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 104(12), 639-659.
- Wimsatt, W. C. (1972). Teleology and the logical structure of function statements. *Studies in the History and Philosophy of Science*, 3, 1-80.
- Wouters, A. G. (2003). Four notions of biological function. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 34(4), 633-668.

CHAPITRE 7

OBJECTIVITÉ SOCIALE ET DÉFINITION DU TROUBLE MENTAL : PROPOSITION DE CRÉATION D'UN PANEL AU SEIN DU DSM

Dans les chapitres précédents, j'ai argumenté que l'idéal de la science indépendante des valeurs et la quête d'une notion de dysfonction neutre devaient être abandonnés. Or un nouvel idéal doit être proposé pour parvenir à répondre aux problèmes soulevés par la définition du trouble mental. J'ai soutenu que cet idéal est celui de l'objectivité sociale (OS), comme proposée par Longino, qui permet à la fois d'accommoder la présence de valeurs en psychiatrie ainsi que d'identifier et de critiquer les valeurs qui apparaissent problématiques. Dans le chapitre 5, j'ai suggéré que certaines modifications devaient être apportées à l'OS, principalement en ce qui a trait à l'objectivité empirique et à la nature de la diversité de la communauté épistémique. Ce chapitre m'a conduit à soutenir une version amendée de l'OS, l'OS*. Dans le chapitre 6, j'ai montré comment il serait possible de parler d'un concept empiriquement objectif de trouble mental en empruntant à la théorie de la fonction comme rôle causal mécaniste.

Dans ce chapitre final, je retournerai à la manière dont la définition du trouble mental pourrait être dite socialement objective sur le plan des processus sociaux de production de connaissance. Je vais illustrer la manière dont l'OS* pourrait prendre forme dans le contexte de la définition du trouble mental donnée par l'*Association de psychiatrie américaine* (APA) et comment les différentes propositions d'amendements présentées

au chapitre 5 peuvent y être intégrées. L'OS* est une théorie générale de l'objectivité en science. En ce sens, la manière d'appliquer l'OS* dans des situations spécifiques est sujette à interprétation, et les paramètres exacts de cette application ne sont pas précisés par Longino (ni par qui que ce soit d'autre, d'ailleurs). Ce chapitre se veut donc exploratoire, et ses buts modestes : il s'agit d'offrir une ébauche de ce à quoi pourrait ressembler une définition socialement objective du trouble mental dans le DSM, sans toutefois prétendre que ce soit la seule manière de l'appliquer. Dans la section 7.1, je vais exposer les processus qui ont guidé la révision du DSM-5. Dans la section 7.2, je vais m'intéresser plus particulièrement à la manière dont la définition du trouble mental a été révisée dans le DSM-5. Dans la section 7.3, je vais analyser les processus de révision de la définition du trouble mental avec le cadre de l'OS* (section 7.3.1), examiner une proposition de création d'un panel qui aurait pour tâche de discuter cette question (section 7.3.2), pour ensuite évaluer cette proposition grâce aux outils de l'OS* (section 7.3.3). Dans la dernière section (7.4), je me pencherai sur une limite potentielle de ce panel, soit l'influence de l'industrie pharmaceutique sur la définition du trouble mental. En conclusion, je discuterai d'autres limites qui pourraient être rencontrées par le panel et des limites de l'OS* dans ce cadre.

7.1 Structures de prise de décision dans le DSM-5

Dans cette section, j'exposerai le processus général de révision du DSM-5 pour mieux comprendre à quoi pourrait ressembler le processus de révision de la définition du trouble mental. Le processus de révision du DSM-5 a commencé en 1999, chapeauté par l'APA, en collaboration avec la World Health Organization (WHO) Division of Mental Health, la World Psychiatric Association (WPA), et le National Institute of

Mental Health (NIMH). À partir de 1999, plusieurs conférences de planification ont été tenues, et un groupe de travail destiné à fixer les orientations de recherche du futur DSM a été créé. Ce groupe de travail a produit des « livres blancs » (*white papers*) destinés à identifier les axes de recherche et de travail pour réviser le DSM en vue de la future publication du DSM-5. Le tout a été rassemblé et publié dans une monographie intitulée *A Research Agenda for DSM-V* (Kupfer *et al.*, 2002). Les différents chapitres de l'ouvrage se voulaient des recensions des écrits visant à cartographier les enjeux liés à différentes problématiques complexes, dans le but de guider les prises de décisions futures. Ce sont au total six groupes de travail qui se sont penchés sur de grands thèmes de recherche à venir : les enjeux de nomenclature, les neurosciences et la génétique, les diagnostics neurodéveloppementaux, les troubles de personnalité et relationnels, le lien entre les troubles mentaux et les handicaps, et les enjeux de psychiatrie transculturels (Kupfer *et al.*, 2002; Regier *et al.*, 2002). Par la suite, trois livres blancs ont été publiés en rapport avec d'autres enjeux considérés importants : le lien entre le genre et les troubles mentaux, les diagnostics dans les populations âgées et les troubles mentaux chez les nourrissons et les jeunes enfants.

Après la publication de recommandations émises par ces groupes de travail et la tenue de plusieurs autres conférences et assemblées destinées à la planification (2004-2008), les président et vice-président responsables de la rédaction du DSM-5 ont été nommés en 2006, respectivement David Kupfer et Darrel A. Regier. Les membres du groupe de travail général du DSM [*Task Force*] ont également été désignés, en plus de la création de 13 groupes de travail spécifiques [*work groups*] responsables de se pencher sur des ensembles de catégories diagnostiques spécifiques du DSM (p. ex. : troubles développementaux, troubles de la personnalité, etc.) et des enjeux particuliers (p. ex. : psychiatrie transculturelle, enjeux de genre). Les groupes de travail spécifiques comprenaient entre 8 et 15 membres d'expertise multidisciplinaire (quoique majoritairement composés de psychiatres, APA, 2013, p. 6), tous désignés par le conseil d'administration de l'APA. Entre 2008 et 2012, chacun des groupes de travail

était responsable d'effectuer une recension des écrits et une méta-analyse reliées aux propositions de modifications concernant les diagnostics particuliers du DSM. En complément au travail de ces groupes, plus de 400 membres sans droit de vote ont été nommés comme conseillers dans les divers groupes de travail et ont également participé au processus de révision. Le groupe de travail général [*Task Force*] était quant à lui responsable des enjeux plus généraux comme l'organisation générale du manuel (Regier *et al.*, 2011; Browne, 2017). Entre 2010 et 2012, les ébauches de modification des critères diagnostics ont été publiées en ligne (dsm5.org) et commentées par le public, pour une période de deux mois. Les commentaires ont été par la suite passés en revue de manière systématique et intégrés par les groupes de travail, lorsque jugés pertinents. Qui plus est, plusieurs pétitions se prononçant contre ou en faveur de certaines modifications proposées ont été soumises à l'APA (APA, 2013, p. 7).

Des essais de terrains [*field trials*] ont débuté en 2010 dans le but de démontrer la fiabilité, la validité, la faisabilité et l'utilité clinique des changements proposés aux diagnostics. En parallèle, le *Scientific Review Committee* (SRC) a été créé. Le SRC avait pour tâche de réviser les propositions de modifications émises par les groupes de travail du DSM-5 selon les éléments de preuve disponibles dans les écrits. Un autre comité, le *Clinical and Public Health Committee* (CPHC), a également été créé pour évaluer les impacts des propositions de modifications sur les domaines clinique et de santé publique. En 2012, chacun des groupes de travail a soumis l'ébauche finale des modifications proposées au SRC, lequel a émis ses commentaires. Finalement, les propositions de modifications ont été examinées par un « groupe au sommet » [*Summit Group*], dont faisaient partie des membres des groupes de travail, du SRC et du CPHC, ainsi que d'un groupe de révision en psychiatrie légale. Le groupe au sommet était également composé d'experts externes et de membres du comité exécutif du CA de l'APA (Yager et McIntyre, 2014). Les recommandations du groupe au sommet ont ensuite été transmises à l'assemblée de l'APA et à son CA (APA, 2013). Après

plusieurs étapes de discussions, de recommandations et de révisions impliquant des experts internes et externes, le manuel a finalement été publié en 2013.

Le processus de révision du DSM-5 fait donc appel à plusieurs structures et se déroule selon plusieurs étapes. Il faut noter que plusieurs modifications ont été apportées au processus de révision du DSM-5 par rapport aux révisions précédentes. D'abord, le fait que les propositions de modifications aient été publiées en ligne, ouvrant ainsi un espace pour que le public commente ces modifications, a été considéré par plusieurs comme une nette amélioration vis-à-vis le précédent processus de révision du DSM-IV (p. ex. : Stein et Phillips, 2013; Bueter, 2019; pour une critique, voir Spitzer, 2009). Par ailleurs, la perspective des patients a été davantage intégrée au cours du processus de révision (ce qui avait été critiqué par Sadler et Fulford, 2004). Le lancement du site web a été précédé d'une consultation auprès de représentants de groupes de patients et des membres de leur famille, tels que le *National Alliance on Mental Illness* (NAMI), la *Depression and Bipolar Support Alliance* (DNSA), and *Mental Health America* (MHA). Cette consultation visait à cibler et à discuter les enjeux qui affectent directement les patients, tels que le stigma et l'accès aux soins et à l'assurance (etc.) (Regier *et al.*, 2010). Les patients pouvaient aussi exprimer leurs perspectives via les commentaires en ligne. Par exemple, Stein et Philipps (2013) rapportent que le groupe de travail ayant pour tâche d'analyser les propositions de modifications des troubles obsessionnels compulsifs a pris en compte les commentaires des patients émis sur le site web du DSM-5. En guise d'illustration, des débats existaient dans le groupe de travail quant à la création des nouveaux diagnostics de l'acné excoriée (ou dermatillomanie), un trouble qui se caractérise par le grattage compulsif de la peau (principalement de ses imperfections) et le trouble d'accumulation compulsif (ou syllogomanie), soit la tendance à accumuler de manière compulsive des objets, indépendamment de leur utilité. Bien que des éléments de preuve existent en faveur de la création de ces catégories, le fait que de nombreux patients se soient exprimés pour la création de ces diagnostics spécifiques a contribué à la décision du groupe de travail d'ajouter ce

diagnostic au DSM-5. Qui plus est, au sein des membres de la *Task Force* du DSM-5 se trouvait un patient nommé pour représenter des intérêts des patients. Son expertise a été reconnue pour pointer l'impact des diagnostics psychiatriques sur l'accès au soin, sur la perception culturelle et sociale d'un diagnostic psychiatrique et sur la possibilité de diagnostics inadéquats (p. ex. : faux positifs ou faux négatifs) (Regier *et al.*, 2010). Par ailleurs, le DSM-5 écrit :

Many health professions and educational groups have been involved in the development and testing of DSM-5, including physicians, psychologists, social workers, nurses, counselors, epidemiologists, statisticians, neuroscientists, and neuropsychologists. Finally, patients, families, lawyers, consumer organizations, and advocacy groups have all participated in revising DSM-5 by providing feedback on the mental disorders described in this volume. Their monitoring of the descriptions and explanatory text is essential to improve understanding, reduce stigma, and advance the treatment and eventual cures for these conditions (APA, 2013, p. 6).

Le processus de révision du DSM a donc été beaucoup plus inclusif que les versions précédentes, surtout en ce qui concerne l'inclusion des patients dans le processus.

7.2 Le processus de définition du trouble mental dans le DSM-5

Dans la section précédente, j'ai exposé les processus de prise de décision concernant les modifications générales proposées pour le DSM-5. Dans cette section, je vais d'abord m'intéresser brièvement aux structures de prise de décision concernant la définition de la santé mentale dans les premières éditions du DSM, pour ensuite me pencher plus longuement sur les structures de prise de décision du DSM-5. Cette

présentation a pour but de me donner les bases pour appliquer l'OS* à la définition du trouble mental du DSM dans la section suivante. Pour reconstruire la manière dont les différentes définitions du trouble mental du DSM ont été élaborées jusqu'au DSM-5, je me baserai principalement sur l'analyse de Cooper (2015; voir aussi Demazeux, 2021 et Kinghorn, 2013).

Rappelons qu'au début des années 1970, la définition du trouble mental était perçue comme importante pour la psychiatrie américaine en raison des controverses dans lesquelles elle était impliquée, soit la crise de l'homosexualité et le mouvement antipsychiatrique. Dans l'introduction, j'ai passé en revue ces débats ainsi que les différentes versions de définitions proposées. Ici, je mets en évidence la manière dont les décisions ont été prises, enjeu qui a attiré beaucoup moins d'attention dans les écrits. Dans la deuxième moitié des années 1970, l'APA a mis en place un groupe de travail pour définir « maladie mentale » (*mental illness*) et « psychiatres ». Or il semble que ce groupe de travail n'ait finalement pas influencé la conception du DSM-III, mais que leur travail ait résulté en la publication par l'APA d'un dépliant intitulé « What Is a Psychiatrist? » en 1978 (Cooper, 2015). En parallèle, en 1975, le processus de révision du DSM-III a débuté, avec entre autres la nomination de Robert Spitzer comme président de la Task Force. En 1976, une première ébauche de définition du trouble mental a été proposée par Spitzer et Endicott pour être incluse dans le DSM (basée sur Spitzer et Wilson, 1975; voir Demazeux, 2021). Cette ébauche a été soumise à l'assemblée annuelle de l'APA et perçue négativement par ses membres. En 1978, une deuxième tentative de définition a été proposée à l'APA par Spitzer et Endicott (et publiée dans Spitzer et Endicott, 1978), dans laquelle le trouble mental était défini comme une classe particulière de trouble médical. Cette fois, la définition a été soumise directement à la Task Force et a été à nouveau rejetée. Quelques mois avant la publication du DSM-III (1980), une dernière ébauche a été proposée, rédigée par Spitzer et d'autres membres de la Task Force. Cette version a été incluse dans la troisième édition du DSM (voir aussi Kinghorn, 2013). Par la suite, dans le DSM-III-

R, malgré le fait que très peu de discussions à propos de la définition du trouble mental aient eu lieu, la définition a été modifiée. Les modifications apportées à la définition ont été dirigées à nouveau par Spitzer, toujours président de la Task Force. En ce qui concerne le DSM-IV, cette fois dirigée par Frances, le processus de révision a débuté aux alentours de 1988. Au final, la définition du trouble mental donnée dans le DSM-III-R y a simplement été reprise, avec l'ajout du passage suivant :

Although this manual provides a classification of mental disorders, it must be admitted that no definition adequately specifies precise boundaries for the concept of 'mental disorder.' The concept of mental disorder, like many other concepts in medicine and science, lacks a consistent operational definition that covers all situations. All medical conditions are defined on various levels of abstraction – for example, structural pathology (e.g., ulcerative colitis), symptom presentation (e.g., migraine), deviance from a physiological norm (e.g., hypertension), and etiology (e.g., pneumococcal pneumonia). Mental disorders have also been defined by a variety of concepts (e.g., distress, dysfunction, dyscontrol, disadvantage, disability, inflexibility, irrationality, syndromal pattern, etiology, and statistical deviation). Each is a useful indicator for a mental disorder, but none is equivalent to the concept, and different situations call for different definitions.

Despite these caveats, the definition of *mental disorder* that was included in *DSM-III* and *DSM-III-R* is presented here because it is as useful as any other available definition and has helped to guide decisions regarding which conditions on the boundary between normality and pathology should be included in DSM-IV (APA, 1994, p. xxi).

Cette citation révèle que l'attitude vis-à-vis la définition du trouble mental a changé. En effet, Cooper (2015) rapporte qu'aucun groupe de travail ne s'est penché sur la définition du trouble mental, et le nouveau président de la Task Force, Allen Frances, n'a montré que peu d'intérêt pour cette question (pour la réitération de la position de Frances dans le contexte actuel, voir Widiger et Frances, 2012). Il en va de même pour le DSM-IV-R. Cooper explique que sur cette question, les présidents de la Task Force avaient un grand pouvoir discrétionnaire. Robert Spitzer percevait la définition du

trouble mental comme utile et y a consacré beaucoup d'énergie, alors qu'Allen Frances, président de la Task Force du DSM-IV et du DSM-IV-R, n'avait que peu d'intérêt pour la question et n'y a pas consacré de ressources. Il faut dire, comme le note Cooper, que les tensions politiques qui avaient visé l'APA dans les années 1970 et au début des années 1980 et qui semblaient exiger qu'on définisse ce que l'on entendait par « trouble mental » s'étaient apaisées. Définir celui-ci apparaissait dès lors beaucoup moins urgent et primordial (Cooper, 2015, p. 88-89).

Comme je l'ai dit plus tôt, le processus de révision du DSM-5 s'est déroulé sur presque 15 ans (de 1999 à 2013, année de parution du manuel). Cette période a vu la résurgence de l'intérêt dans la question de la définition du concept de trouble mental. Par exemple, dans l'ouvrage de Kupfer et coll. (2002), un chapitre intitulé *Basic Nomenclature Issues for DSM-V* aborde la définition du trouble mental comme l'un des enjeux centraux de nomenclature pour le DSM-5 à venir (voir Kupfer *et al.*, 2002, p. xix-xxi, pour le processus de rédaction de l'ouvrage général). Dans ce chapitre on peut lire :

Despite the difficulties involved, it is desirable that DSM-V should, if at all possible, include a definition of *mental disorder* that can be used as a criterion for assessing potential candidates for inclusion in the classification, and deletions from it. If for no other reason, this is important because of rising public concern about what is sometimes seen as the progressive medicalization of all problem behaviors and relationships. (Rounsaville *et al.*, 2002, p. 3)

La définition du trouble mental a donc été perçue par ce groupe de travail comme une tâche importante en raison des controverses entourant la médicalisation en psychiatrie. La section du chapitre sur la définition du trouble mental présente ensuite différentes approches pour définir le trouble mental et conclut avec des recommandations de recherches futures. De manière générale, le groupe de travail propose que des recherches soient menées pour connaître de quelle manière les cliniciens conceptualisent le trouble mental et quelles sont les implications d'un choix de

définition par rapport à un autre. L'agenda de recherche se construit selon trois recommandations principales, résumées comme suit (Rounsaville *et al.*, 2002, p. 6-7) :

- 1) Analyser le concept de trouble mental tel qu'il apparaît dans le DSM, et évaluer la manière dont les différents diagnostics se conforment à la conceptualisation générale;
- 2) Réaliser un sondage aux États-Unis et à l'international pour déterminer la ou les significations données au concept de trouble mental par les professionnels de la santé mentale, les patients, et des membres de groupes sociaux et ethniques différents;
- 3) Conduire des recherches au sein des populations susmentionnées pour déterminer la distinction entre les personnes diagnostiquées avec des troubles mentaux et les personnes présentant des symptômes bénins associés à ces diagnostics.

Les recommandations quant à la définition du trouble mental visaient donc principalement à obtenir un portrait de la manière dont les personnes concernées par la définition du trouble mental ou usant de ce concept le comprenaient.

Or après la publication de ces recommandations, aucun groupe de travail n'a été créé pour se pencher sur les enjeux de nomenclature, dont la définition du trouble mental faisait partie. La Task Force s'est tout de même penchée sur la question, mais elle n'a pas suivi les recommandations de recherche de Rounsaville *et al.* Des ébauches de modification de définition ont été publiées sur le site internet du DSM-5 pour être commentées par le public. Cependant, les commentaires publiés en ligne ne sont pas accessibles, car ils ont été retirés, et il n'est pas possible de savoir si et comment les commentaires ont été pris en compte par rapport à la définition du trouble mental. Grâce à son travail de recherche, Cooper (2015) rapporte que deux versions de la définition du trouble mental ont ultimement été débattues : une version révisée du

DSM-IV-TR, présentée par Dan J. Stein, Katharine A. Phillips, Derek Bolton, K.W.M Fulford, John Z. Sadler, et Kenneth S. Kendler, ainsi qu'une version proposée par le *Impairment and Disability Assessment Study Group*, l'un des groupes de travail du DSM-5. Ce groupe avait pour tâche de s'interroger sur les notions de détresse, de préjudice et de handicap, et leurs liens avec le trouble mental. Un point crucial séparait les deux versions : la présence ou l'absence du critère de « détresse ou de handicap » dans la définition. La définition de Stein *et al.* (2010) conservait le critère de préjudice et modifiait la formulation du DSM-IV pour la rendre plus cohérente et moins redondante, alors que la définition du *Impairment and Disability Assessment Study Group* proposait de retirer le critère de préjudice (voir Narrow et Kuhl, 2010 pour une justification de cette position). Comme le note Cooper, lorsque l'on regarde la définition publiée par le DSM-5, il semble qu'un compromis ait été fait entre les deux versions, puisque l'on considère que le trouble mental est « *généralement associé à une importante détresse ou un handicap dans les activités sociales, professionnelles, ou d'autres activités?? importantes* » (APA, 2013, les italiques sont les miennes)¹.

Il est possible de tirer quelques conclusions par rapport à la manière dont la définition a été révisée. L'analyse de Cooper et les textes publiés de l'APA révèlent que depuis le DSM-IV, très peu de débats officiels ont visé la définition du trouble mental (mis à part, pour le DSM-5, la publication des articles de Stein *et al.*, de Narrow et Kuhl, et le chapitre de Rounsaville *et al.*). Cooper écrit :

One of the most striking facts about the changes that have been made to the conceptualisation of the link between disorder and harm in the DSMs is how little thought has gone into these changes and how little debate they have attracted. The DSM-IV archives reveal that no group worked on the definition

¹ Ici, une limite importante de mon analyse doit être notée : nous n'avons pas accès aux raisons exactes pour lesquelles c'est cette définition qui a été choisie, ni à la manière dont cette décision a été prise. En effet, ces informations n'ont pas été rendues publiques par l'APA. L'hypothèse d'un compromis émise par Cooper me semble cependant plausible dans le contexte.

of disorder, or on the role that the requirement that disorder cause distress or impairment might play in the classification. Jerome Wakefield was approached, but only in late 1992, when most of the work on the DSM-IV was already complete. Similarly, with the DSM-5, competing definitions of disorder remained in drafts until after the rest of the classification was fairly much complete (Cooper, 2015, p. 93).

Ainsi, bien que la définition ait de nouveau paru à plusieurs comme un enjeu important du DSM-5, aucune structure n'a réellement été mise en place pour assurer que la question soit traitée de façon adéquate. En effet, aucun groupe de travail ne s'est penché sur cette question durant le processus de révision du DSM (bien que des groupes de travail aient été créés pour aborder d'autres enjeux comme le genre ainsi que la psychiatrie transculturelle, et que le groupe de travail sur le handicap se soit finalement penché sur la question du lien entre « handicap » et « trouble mental »).

Un autre point sur lequel il convient d'insister est l'impact réel de la définition du trouble mental sur le reste du DSM. Mis à part les définitions des DSM-III et DSM-III-R et leur lien avec l'exclusion de l'homosexualité, la définition du trouble mental n'a eu que peu d'impact sur le reste du contenu du DSM (p. ex. : Kinghorn, 2013). Même Spitzer et Williams (1982, cité dans Kinghorn, 2013, p. 52), les rédacteurs de cette première définition, ont reconnu qu'elle avait eu peu d'influence sur les décisions au sujet des autres éléments du DSM-III. En effet, dans le cas du DSM-III, la définition a été rédigée après que la majeure partie des modifications par rapport au DSM-II aient été faites. Dans le cas du DSM-IV, la définition semble ne pas avoir joué de rôle pour départager des enjeux controversés par rapport à certaines catégories diagnostiques spécifiques (voir Frances dans Philips *et al.* 2012). Dans le DSM-5, même si la définition a de nouveau connu une vague d'intérêt et que celle-ci s'est vue modifiée, ces modifications ne semblent pas avoir réellement affecté ni la structure ni le contenu du manuel (Amoretti et Lalumera, 2019; Cooper, 2015; First et Wakefield, 2013; Kinghorn, 2013) puisqu'encore une fois, elle a été produite beaucoup trop tard pour

pouvoir influencer les modifications du DSM-5. Cooper cite les propos de Regier, vice-président de la Task Force du DSM-5 qui dit : « almost none of the DSM-5 Task Force or Work Group meetings struggled with these definitional issues as they evaluated the research literature to determine the evidential basis for revisions (Regier, 2012, p. 292-293, cité dans Cooper, 2015, p. 92). Ainsi, parce que la définition du trouble mental a été discutée très tard dans le processus de révision du DSM-5, les modifications apportées n'ont pas pu influencer les révisions des catégories diagnostiques particulières. Cela va à l'encontre des recommandations de Rounsaville et coll., qui plaidaient pour que la définition du trouble mental puisse servir à apaiser les tensions entourant les controverses autour de la médicalisation par la psychiatrie. Ainsi, bien que la définition du trouble mental ait été vue comme un moyen de se pencher sur l'enjeu de la médicalisation et de guider les décisions par rapport à la création ou la révision de catégories diagnostiques particulières, elle n'a finalement jamais vraiment joué ce rôle.

7.3 La définition du trouble mental et les standards de l'objectivité sociale

Dans la section précédente, j'ai présenté les différentes structures de prise de décision par lesquelles les révisions à la fois du contenu (critères diagnostics ainsi que nouveaux troubles ou nouvelle organisation du contenu) et de certains des concepts principaux (comme ceux de trouble mental) du DSM ont été effectuées. Alors que la section 7.1 a montré que le DSM-5 a modifié ses structures pour devenir plus inclusif et transparent lors de son processus de révision, la section 7.2 a mis en lumière la quasi-absence de débat pour traiter la question spécifique de la définition du trouble mental. Dans cette section, j'évaluerai si et dans quelle mesure le processus de révision du DSM-5 est

conforme aux exigences de l'OS* quant à la définition du trouble mental. Dans la section 7.3.1, je présenterai une première analyse critique de ce processus de révision avec les outils de l'OS*. En me basant sur les résultats de cette analyse, je vais suggérer que le DSM-5 ne respecte pas les normes de l'OS* en ce qui concerne la définition du trouble mental. Ensuite, dans la section 7.3.2, je présenterai une proposition de réforme du DSM élaborée par Browne (2017). Malgré le fait que cette proposition ne soit pas ancrée dans le cadre de l'OS*, elle me semble être compatible avec celle-ci, moyennant certains amendements, ce que je démontrerai dans la section 7.3.3. En m'appuyant sur la proposition de Browne, je défendrai donc qu'un comité composé de membres représentant différents groupes sociaux concernés par la définition du trouble mental doit être créé, comité dont l'une des tâches serait justement de fournir une description de la notion de trouble mental².

7.3.1 Critique des processus de prises de décisions quant à la définition du trouble mental grâce au cadre de l'objectivité sociale

Dans un premier temps, rappelons les caractéristiques de l'OS* défendues par Longino et ma proposition d'amendement quant à la diversité de la communauté épistémique. De manière générale, l'OS* prescrit une communauté épistémique qui serait soumise

² Il est possible de penser à d'autres tâches que pourrait accomplir le panel. Par exemple, le panel pourrait identifier quelle est la conception dominante du trouble mental, voir si elle est identique chez tous/toutes, voir quels sont les effets de celle-ci sur le reste du DSM, etc. (en accord avec les recommandations du *white paper* de 2002 rédigé par Rounsaville *et al.*). Le panel pourrait aussi avoir pour fonction de s'assurer qu'une fois la définition choisie, les différents diagnostics de troubles mentaux du DSM s'arriment à cette définition générale. On s'assurerait alors que la définition joue véritablement son rôle de démarcation pour contrer l'absence de pouvoir ou de normativité des précédentes définitions du DSM.

à un dialogue critique provenant de perspectives diversifiées, menant à une critique intersubjective transformatrice. Les normes précises de l'OS* sont au nombre de quatre :

- 1) Des lieux de discussions critiques accessibles;
- 2) Une réceptivité à la critique de la part de la communauté épistémique;
- 3) Des standards publiquement reconnus;
- 4) Une égalité d'autorité intellectuelle « tempérée » [*tempered equality of intellectual authority*] au sein de la communauté épistémique.

Ces quatre normes ont pour but de réguler la présence de valeurs sociales et culturelles en science en venant contraindre les structures de production de connaissance. Dans le chapitre 5, en me basant sur la théorie féministe du *standpoint*, j'ai proposé que la diversité de la communauté épistémique se comprenne en fonction de la diversité des *standpoints*, et donc qu'elle ne soit pas simplement une diversité de points de vue quelconques³. Cet amendement ouvre la porte à des communautés de recherche qui intégreraient des membres externes à la communauté scientifique lorsque leur perspective est pertinente pour identifier les biais et valeurs dans la recherche menée. Les normes de l'OS* étant de nature très générale et pouvant prendre plusieurs formes selon les contextes d'investigation scientifique (p. ex. : Biddle, 2007), il semble nécessaire d'en adapter la forme à notre type d'entreprise. Qui plus est, le respect de ces normes s'évalue par degré. Ainsi, une communauté épistémique peut respecter plus ou moins les normes de l'OS*. Il convient donc de voir comment les structures de prise de décision de l'APA par rapport à la définition du trouble mental du DSM se rapprochent ou s'écartent des normes générales de l'OS*. Avant de poursuivre, il

³ Rappelons que l'idée de *standpoint* réfère à un point de vue informé par la situation sociale dans laquelle l'individu se trouve et pour laquelle il a une conscience de sa situation, par exemple, être un patient et avoir conscience de sa situation et de la relation de pouvoir qu'elle implique avec le corps médical.

convient de noter une difficulté à laquelle fait face cette analyse : l'opacité des structures de prise de décision quant à la définition du trouble mental au sein de l'APA. Je procéderai donc à partir des informations accessibles, mais en raison du manque d'accès à des informations plus détaillées, la portée de mon analyse demeurera limitée.

D'abord, examinons si le processus respecte les normes 1 et 2. En ce qui concerne le processus de révision de la définition du trouble mental, y avait-il des lieux de discussions et de critiques accessibles, et cette critique a-t-elle été intégrée convenablement? Comme exposé dans la section 7.2, deux espaces de discussions et de critiques peuvent être identifiés : le groupe de travail responsable de la publication d'un livre blanc sur les enjeux de nomenclature du DSM et la publication des propositions de révision de la définition sur le site web du DSM-5 avec la possibilité de commentaires par le public. La formation du groupe de travail qui s'est penché sur la question de la définition du trouble mental et qui a publié un chapitre dans l'ouvrage de Kupfer et coll. (2002) respecte la norme 1. En effet, un espace de discussion a été créé pour traiter la question. Or comme je l'ai dit, les recommandations émises par le groupe n'ont pas été prises en compte. En ce sens, le processus de révision du DSM ne respecte pas la norme 2, soit la réceptivité et l'intégration de la critique, puisqu'il n'y avait pas de mécanismes en place pour assurer que les recommandations seront suivies. Ces recommandations étaient consultatives uniquement. Quant aux ébauches de définition qui ont été publiées sur le site web du DSM-5 pour être commentées, elles représentent également un pas dans la direction de la norme 1. Or, les différentes ébauches et les commentaires ont été retirés du site et sont désormais inaccessibles. Il n'est pas non plus possible de savoir si les commentaires ont été pris en compte pour décider quelle définition allait être publiée dans le DSM-5. Il est donc difficile de se prononcer sur le respect de la norme 2 à ce moment précis du processus de révision. Or ce qui semble plus clair, c'est que les commentaires émis sur le site du DSM-5 étaient uniquement consultatifs. À cet effet, il n'y a pas de structures qui permettent de

s'assurer que la critique émise à travers les commentaires en ligne ait bel et bien été intégrée.

Par ailleurs, l'opacité du processus quant à la définition du trouble mental me pousse à penser que la troisième norme de l'OS*, soit celle concernant le fait qu'il doive exister des standards publiquement reconnus de prises de décision, n'est pas remplie. En effet, les raisons pour lesquelles la définition publiée dans le DSM a été choisie n'ont pas été rendues publiques, ce qui contrevient à l'idéal de transparence promu par la norme 3. Les commentaires sur la définition du trouble mental publiés en ligne n'étant plus publics, le processus contrevient là encore à la norme 3. Finalement, la norme 4, celle concernant l'égalité d'autorité intellectuelle tempérée, me semble plus difficile à appliquer pour la question spécifique de la définition du trouble mental. Rappelons que cette norme demande à ce que l'appartenance à un groupe social subordonné n'affecte pas l'expertise d'un individu sur un sujet donné. Étant donné que la définition du trouble mental a été très peu discutée et que très peu de structures officielles ont été créées, en plus de l'absence de transparence par rapport aux prises de décisions à ce sujet, il y a peu de moyens pour juger la manière dont la quatrième norme a été respectée (si elle a été respectée). Il n'y a pas réellement d'éléments qui pointent en direction de discriminations *explicites* dans le processus de révision de la définition (mais il n'y a pas non plus d'éléments qui pointent dans la direction inverse). Or tant que des structures plus formelles et plus transparentes ne seront pas créées, il y a un risque élevé que la norme 4 ne soit pas respectée. Qui plus est, en ce qui concerne la diversité de *standpoints* de la communauté épistémique, aucune structure n'assure le dialogue critique intégrant divers *standpoints* dans le processus de révision du DSM de la définition du trouble mental. Les commentaires en ligne se font sur une base volontaire. En ce sens, il n'est pas garanti que les diverses perspectives pertinentes aient exprimé leur voix sur la question. La composition du groupe de travail responsable du chapitre sur les enjeux de nomenclature n'était pas non plus très diversifiée (tous des psychiatres, bien que certains se soient spécialisés sur des enjeux

plus conceptuels, comme Renato D. Alarcón sur les enjeux transculturels en psychiatrie, voir par exemple Alarcón, 2014). Le processus de révision du DSM-5 par rapport à la définition du trouble mental ne respecte donc pas la nature de la diversité de l'OS* en termes de *standpoints*.

7.3.2 L'apport de T.K. Browne : un panel d'éthique responsable des valeurs qui influencent la catégorisation psychiatrique pour le DSM-5

L'analyse proposée dans la section précédente se montre plutôt critique par rapport au processus de révision de la définition du trouble mental du DSM-5. En effet, dans son état actuel, ce processus ne répondrait pas suffisamment aux normes de l'OS*. Or, une approche plus constructive a été récemment proposée par Tamara K. Browne pour penser la révision des catégories diagnostiques du DSM (2017a, 2017b; pour une approche similaire, voir Sisti et Johnson, 2015), mais en dehors du cadre de l'OS*. Le point de départ de Browne est la présence de controverses importantes autour de la médicalisation de certaines conditions par le DSM. Elle suggère la création d'un panel d'éthique qui viendrait compléter le processus de révision du DSM menant à l'inclusion ou à l'exclusion de nouvelles catégories diagnostiques dans le manuel. Selon Browne, les controverses auxquelles fait face le DSM peuvent être expliquées par deux facteurs majeurs : d'abord, certaines décisions ne relèvent pas uniquement des faits, mais aussi des jugements de valeur. Elle cite en exemple le trouble de la dépression majeure, décrit dans le DSM selon des critères internes à l'individu (p. ex. : changement de poids, troubles du sommeil, fatigue, pensées suicidaires, etc.). Or il serait possible de décrire ce trouble en rapport avec le contexte dans lequel se trouve l'individu (p. ex. : milieu de travail ou relations sociales problématiques). Le fait de ne

pas considérer le contexte dans lequel se trouve l'individu, mais seulement des critères internes à l'individu pour le diagnostic de la dépression majeure n'est pas basé uniquement sur une lecture des données empiriques, mais est motivé par des décisions (voir aussi Sadler, 2005 pour une revue plus complète des types de jugement de valeur que l'on retrouve dans la classification psychiatrique). Ces jugements de valeur qui expliquent certaines décisions dans le processus de révision du DSM peuvent être la source de controverses. Deuxièmement, sans nier l'importance des données empiriques dans le processus de révision du DSM (voir la section 7.1), Browne soutient que l'impact pratique du DSM sur différentes sphères est majeur (p. ex. : le domaine légal, les politiques publiques en santé mentale, l'économie de la santé, l'accès au soin, les représentations de la santé mentale au sein du public de manière générale, etc.). Les controverses créées par la publication du DSM s'expliquent en partie par les effets que les modifications apportées au manuel entraînent dans ces différents domaines. Malgré le rôle des jugements de valeur dans les décisions de révisions du DSM et les impacts pratiques de ces révisions, pour le moment, le DSM n'a pas les moyens de remédier à ces problèmes. Selon Browne, c'est parce que l'APA n'a pas mis en place de structures pour rendre explicites les jugements de valeur qui motivent les décisions. La création d'un panel pourrait apaiser les tensions liées à la médicalisation entraînées par le DSM. Bien que la proposition de Browne vise l'inclusion ou l'exclusion de catégories diagnostiques particulières au sein de DSM, son modèle pourrait s'appliquer à la question de la définition du trouble mental. En effet, si l'on pense comme Rounsaville et coll. 2002 qu'une définition du trouble mental devrait venir guider les décisions par rapport à l'inclusion ou l'exclusion de diagnostics dans le DSM, le panel de Browne pourrait aussi servir à définir le trouble mental.

Expliquons d'abord la position de Browne. Selon elle, les valeurs n'ont pas été rendues explicites ni examinées de manière critique dans le processus de révision du DSM-5. Pour faire face à ce problème, elle recommande donc l'ajout de mécanismes structurels qui permettraient l'analyse critique des valeurs en jeu dans la classification

psychiatrique. La proposition de Browne est similaire à une prise de position de Kendler et coll. (2008; voir aussi Sadler, 2002) par rapport à l'urgence d'aborder les enjeux conceptuels sous-jacents au DSM-5 (la proposition de Browne est cependant plus détaillée). En effet, Kendler et coll. ont défendu la création d'un groupe de travail sur les enjeux conceptuels entourant le DSM, incluant la définition du trouble mental. Bien que la proposition de Kendler et coll. touche les enjeux conceptuels du DSM de manière générale et la proposition de Browne les enjeux de valeurs, leurs deux perspectives semblent pointer dans la même direction : le besoin d'aborder des présuppositions implicites dans la classification du DSM par des *mécanismes structurels*.

Selon Browne, les mécanismes structurels pourraient prendre la forme d'un panel d'éthique indépendant composé d'experts. Le comité serait composé de spécialistes en bioéthique, de sociologues et de philosophes de la psychiatrie. Pour Browne, ce choix de spécialistes s'explique par l'expertise de ce type de chercheurs sur la question des jugements de valeur en rapport avec les classifications en psychiatrie et par les impacts de certaines décisions sur différents groupes de la société (Browne, 2017a, p. 192). Qui plus est, Browne défend que pour que le panel soit réellement indépendant, ses membres ne doivent pas être liés à l'APA, ni avoir de conflits d'intérêts avec l'industrie pharmaceutique. Le financement du panel devrait aussi provenir d'institutions publiques plutôt que de compagnies privées afin d'éviter les conflits d'intérêts potentiels. La tâche du panel serait d'aborder les enjeux liés à la présence de jugements de valeur dans le processus de révision du DSM. Ultimement, le panel aurait quatre rôles : 1) rendre explicites les valeurs influençant les révisions proposées, 2) critiquer les valeurs en jeu dans les révisions proposées, 3) identifier les effets potentiels des modifications et 4) suggérer des moyens de minimiser les effets potentiellement préjudiciables causés par ces modifications. Ainsi, le panel aurait pour mission de passer en revue les différentes propositions de modifications de classification et d'approuver ou de rejeter chacune d'entre elles. Selon Browne, le panel

devrait avoir droit de veto sur chacune des révisions. Pour parvenir à une décision, Browne recommande de se baser sur le modèle des conférences de consensus entre experts, où les participants discutent de sujets controversés et aboutissent à une recommandation finale (des modèles de conférences de consensus où des non experts sont intégrés à la discussion existent dans les écrits, souvent désignés sous les noms de « modèle danois » ou de « modèle d'interface », voir Lesurtel *et al.*, 2014; Solomon, 2015, chapitre 3). Browne suggère d'ailleurs que le droit de veto accordé au panel soit conditionnel à un vote à majorité simple, par contraste avec le besoin d'aboutir à une décision unanime, caractéristique des modèles classiques de conférence de consensus. Elle juge peu réaliste de parvenir à un consensus au sujet d'enjeux controversés liés aux valeurs⁴. Par ailleurs, Browne conçoit le panel comme une structure devant intervenir à la fin du processus de révision, dans le but d'avoir une vue d'ensemble sur les modifications proposées (Browne, 2017a, p. 191). De plus, elle note que la transparence du panel devrait être préoccupation centrale de ce genre de comité, à la fois en termes de justification du choix des membres du panel que des décisions rendues. Tous ces éléments devraient donc être rendus publics (p. ex. : par la publication de résumés de discussion, etc.) (Browne, 2017a, p. 195). Ainsi, même si le panel était composé uniquement d'experts, le public pourrait avoir l'information requise pour comprendre les recommandations du panel et exprimer son accord ou son désaccord de manière réfléchie (via, par exemple, des procédures de consultation).

⁴ Pour clarifier, Browne reprend du modèle de conférence de consensus l'idée de négociation, c'est-à-dire que chaque membre présent tente de trouver une solution viable à une controverse, mais préfère la méthode de vote pour que soient décidées les recommandations finales. Qui plus est, il faut noter, suivant Kendler et Solomon (2016), que le modèle de consensus est déjà ce qui guide la prise de décision du processus de révision du DSM, au sens où les différents groupes de travail (Work Groups) du DSM fonctionnent de manière générale selon ce modèle. Ce que Browne propose est l'ajout d'un autre groupe de travail responsable de la question des valeurs au sein du DSM, dont les experts responsables de la prise de décision sur ces questions seraient des philosophes, des bioéthiciens et des sociologues, par contraste avec les experts composant les groupes de travail déjà existants, soit majoritairement des psychiatres et d'autres spécialistes de la santé mentale.

C'est donc en suivant ces différents paramètres que le panel pourrait assurer l'identification et la critique des valeurs qui influencent la classification psychiatrique.

7.3.3 Analyse de la proposition de Browne

Comme noté précédemment, la proposition de Browne ne s'ancre pas dans le cadre théorique de l'OS (elle se base plutôt sur l'existence de comité d'éthique dans d'autres domaines de recherche, Browne, 2017a, p. 191; voir Moon et Khin-Maung-Gyi, 2009; Sobel et Hansen, 2006 pour des arguments en faveur de la création de comités d'éthique en recherche). Cependant, dans les écrits en philosophie de la psychiatrie, le panel d'experts proposé par Browne est ce qui se rapproche le plus d'un modèle compatible avec le nouvel idéal d'objectivité sociale que je propose d'adopter pour la révision de la définition du trouble mental dans le DSM (voir cependant Ross, 2002 pour l'idée selon laquelle l'OS devrait être appliquée au processus de révision du DSM, mais sans illustration concrète). En effet, un tel panel vise précisément à rendre explicite et à critiquer les jugements de valeur qui influencent les modifications apportées au DSM par des mécanismes structurels. Par contre, je ne soutiens pas que ce comité soit la seule incarnation possible de l'objectivité sociale appliquée à la définition du trouble mental. Dans un premier temps, je vais présenter les raisons qui me poussent à penser que ce panel est compatible avec le modèle de l'OS*. Dans un deuxième temps, je vais discuter les possibles limites du modèle à l'aune des normes de l'OS*.

De manière générale, je soutiens que la proposition de réforme de Browne permet de mieux faire face à la présence de valeurs dans le processus de révision du DSM que dans sa version actuelle et qu'il pourrait très bien s'appliquer au cas de la définition du

trouble mental. En effet, le panel répond mieux aux normes de l'OS* que le processus de révision du DSM tel qu'il est. Le comité d'experts proposé par Browne a l'avantage de prendre en compte les processus sociaux de production de connaissances et de reconnaître la présence de valeurs en psychiatrie, plutôt que de les ignorer. De ce point de vue, le panel est compatible avec une vision de la science légitimement influencée par les valeurs, dans laquelle l'OS* s'inscrit. Qui plus est, j'ai déjà défendu que les valeurs qui entrent en jeu dans la définition du trouble mental doivent pouvoir être identifiées et débattues. Ce panel constitue un lieu parfait pour exécuter ce genre de tâche. Afin de défendre l'idée que la proposition de Browne est une incarnation possible de l'OS* en psychiatrie (quoique perfectible), je vais l'évaluer à son tour en fonction des standards de l'OS*⁵. De plus, la diversité promue est celle des *standpoints*. Selon moi, la norme 1 pourrait être atteinte avec l'adoption d'un tel panel. En effet, le panel devient un lieu de critique pour que soient identifiées et critiquées les valeurs présentes, permettant ainsi que le processus de critique intersubjective transformatrice promu par Longino se produise. Qui plus est, étant donné que le panel détiendrait un droit de veto par rapport aux modifications proposées, la deuxième norme qui concerne l'intégration de la critique est respectée. En effet, le droit de veto assure l'intégration des recommandations du panel au sein du DSM, et non pas seulement une tolérance de la part de l'APA envers leur avis. Browne soutient que les justifications de prise de décisions du panel et de la nomination des membres devraient être rendues publiques. Le panel se conforme donc par là à la norme 3, la transparence des standards d'évaluation. Le panel intègre également différentes perspectives, telles la bioéthique, la philosophie de la psychiatrie et la sociologie. En cela, la proposition de Browne fait un pas en direction de l'idéal de diversité de l'OS*. Ainsi, dans ses grandes lignes, ce panel se rapproche de ce qui est prescrit par la théorie de l'OS*.

⁵ Rappelons à nouveau que les quatre normes de l'OS sont : 1) la disponibilité de lieux de critiques, 2) l'intégration de la critique, 3) des standards d'évaluation publique et 4) l'égalité tempérée d'autorité intellectuelle.

Cependant, il est possible de penser à des améliorations à apporter à ce panel d'experts pour qu'il soit davantage conforme aux exigences de l'OS*. Le panel que propose Browne n'inclut pas de mécanismes structurels pour que la quatrième norme de l'OS*, l'égalité d'autorité intellectuelle tempérée, soit respectée. Comme discuté au chapitre 5, les perspectives racistes ou sexistes (ou d'autres formes d'attitudes discriminatoires) ne sont pas compatibles avec le modèle de l'OS*. Il est donc important de faire respecter cette norme en excluant du panel les perspectives qui discriminaient la voix de certains membres en raison de leur appartenance à un groupe social marginalisé. De plus, l'idéal de diversité de *standpoints* pourrait être atteint de manière plus optimale si d'autres perspectives étaient intégrées au panel. L'une des premières voix pertinentes qui viennent à l'esprit pour définir le trouble mental est celle des psychiatres et des chercheurs sur les troubles mentaux. Par souci d'indépendance, Browne avait défendu un panel uniquement composé de « spécialistes des valeurs » et exclu les psychiatres. Or comme ce sont eux qui utilisent le concept dans leur pratique quotidienne, et comme ils possèdent une connaissance approfondie des fondements scientifiques et des manifestations cliniques des troubles mentaux, l'OS* demanderait à ce que leur point de vue soit pris en compte (p. ex. : Blease, 2017; Pouncey et Merz, 2017 pour des arguments en faveur de l'intégration des psychiatres et d'autres spécialistes des troubles mentaux; voir également Douglas, 2003, 2009 pour la responsabilité morale des scientifiques à porter des jugements de valeur). Tout comme Browne le soutient, les bioéthiciens, les philosophes de la psychiatrie et les sociologues ont également des perspectives pertinentes pour la discussion des valeurs, car une partie de leur expertise nécessite l'analyse de la manière dont les valeurs peuvent façonner les représentations et les pratiques ou les ensembles de valeurs qui sont meilleurs que d'autres. Or d'autres approches en sciences humaines sur la psychiatrie pourraient être intégrées, comme l'anthropologie psychiatrique. De plus, comme la psychiatrie a un passé d'abus envers les femmes et les groupes racialisés (voir le chapitre 4), il est raisonnable de recommander que certains experts en sciences humaines soient formés aux études féministes ou aux *race studies*, car ils seraient sans doute plus à même de mettre en

évidence les hypothèses sexistes ou racistes souvent implicites en psychiatrie. Par ailleurs, la prise en compte de la perspective des patients me semble importante. La perspective des patients peut être considérée comme pertinente en raison de leur expérience de la vie avec les troubles mentaux. Ceux-ci sont également susceptibles d'anticiper les effets de la stigmatisation et de l'accès aux soins qu'une définition particulière des troubles mentaux peut avoir sur une autre (voir Bueter, 2018; Cooper, 2017; Porter, 2017 pour les arguments en faveur de l'intégration des patients dans le processus de révision de manière générale). À cet égard, leur point de vue est significatif. Cela étant dit, il convient de noter que la liste des perspectives proposées ne doit pas être considérée comme exhaustive ou définitive. D'autres perspectives pourraient être intégrées si l'on pouvait justifier que leur point de vue maximise le type de diversité de *standpoints* qu'exige l'OS*.

7.4.1 Le panel et l'industrie pharmaceutique

Dans la section précédente, j'ai défendu que le panel proposé par Browne pourrait constituer une bonne incarnation de l'OS* appliquée à la définition du trouble mental, moyennant quelques amendements. Avant de clore ce chapitre, une possible limite à ce panel doit être discutée : l'intrusion possible d'intérêts problématiques dans le panel. L'influence induite de l'industrie pharmaceutique dans le processus de révision de l'APA par le biais du financement et des conflits d'intérêts est une préoccupation récurrente dans les écrits critiques en psychiatrie (p. ex. : Cosgrove *et al.*, 2006; Cosgrove et Krimsky, 2012). Serait-il possible qu'avec la création de ce panel et la nécessité d'une diversité de *standpoints*, les intérêts commerciaux de l'industrie pharmaceutique influencent le processus de définition des troubles mentaux et

conduisent à une définition très inclusive des troubles mentaux pour augmenter les profits, ce qui entraînerait une médicalisation croissante (p. ex. : Moynihan *et al.*, 2013)? Ce questionnement est lié au fait que la médicalisation implique le plus souvent le développement de traitements psychopharmacologiques pour les conditions nouvellement médicalisées. Ce phénomène est parfois appelé « pharmacologisation » (*pharmaceuticalisation*, voir par exemple Bell et Figert, 2012; Tseris 2018; Williams *et al.*, 2011). Une définition plus inclusive des troubles mentaux pourrait mener à une plus grande médicalisation, et par là une plus grande pharmacologisation.

La question de la relation entre l'industrie pharmaceutique et la psychiatrie est très vaste et la portée de ce chapitre ne permet que d'esquisser des lignes de réponse possibles. Une première chose à noter est que l'influence de l'industrie pharmaceutique est le plus souvent indirecte : à l'heure actuelle, l'industrie pharmaceutique ne participe pas directement à définir le trouble mental ni à établir les critères ou les tests diagnostiques pour les troubles mentaux particuliers. Par contre, la grande majorité de la recherche en psychiatrie est financée par l'industrie pharmaceutique (Muschenga *et al.*, 2010). Qui plus est, il a été établi que 69 % des membres des groupes de travail de révision du DSM-5 avaient des liens avec l'industrie pharmaceutique, par le biais de subventions ou de parts d'investissement, de participation payée à des conférences ou à des formations ou par le fait de siéger sur des conseils d'administration (Cosgrove et Krimsky, 2012; voir aussi Cosgrove *et al.*, 2014; Pilecki *et al.*, 2011). Rappelons que les membres de ces groupes de travail sont ceux qui participent à la prise de décisions concernant les modifications à apporter aux diagnostics psychiatriques particuliers ou la création/suppression de catégories particulières. Ces conflits d'intérêts peuvent être vus comme particulièrement problématiques dans les cas où des traitements pharmaceutiques accompagnent quasi nécessairement le diagnostic (comme pour la schizophrénie, voir Cosgrove *et al.*, 2006). Par ailleurs, des études ont montré que les conflits d'intérêts avec l'industrie pouvaient entraîner des « habitudes de pensée pro-industrie » (Lexchin et O'Donovan, 2010; voir aussi Cosgrove *et al.*, 2014), voire

l'expansion de la définition de pathologie particulière ou l'abaissement de seuils diagnostiques (Moynihan *et al.*, 2013). Qui plus est, la possibilité du développement de traitement pharmaceutique est parfois évoquée comme un argument en faveur de la création d'un diagnostic (voir Greene, 2006 en médecine). Considérant les liens qui unissent les membres des groupes de travail du DSM et l'industrie pharmaceutique, la mobilisation de ce type d'argument peut être vue comme préoccupante. Cosgrove et Wheeler (2013a) parlent à cet effet d'une « capture institutionnalisée du cadre épistémique de DSM par l'industrie », au sens où les découvertes de l'industrie pharmaceutique viennent légitimer l'existence de catégories diagnostiques particulières.

Une illustration de ce phénomène est la création du « trouble dysphorique prémenstruel » (TDPM) dans le DSM-5. Ce trouble a été l'objet de vives controverses impliquant les points de vue féministe et neuropsychologique. Plusieurs chercheuses ont soutenu que l'introduction du TDPM dans le DSM marginalisait l'expérience de colère et de labilité émotionnelle des femmes (p. ex. : Caplan, 1995, 2004; Offman et Kleinplatz, 2004; Cosgrove et Wheeler, 2013b), en plus de minimiser les causes externes de la colère des femmes telles que les abus physiques et émotifs et les systèmes d'oppression institutionnels (p. ex. : Caplan, 1995; Caplan, McCurdy-Myers et Gans, 1992; Cosgrove et Caplan, 2004). Qui plus est, la validité de la catégorie a été remise en question (Callaghan *et al.*, 2009; Di Giulio et Riessing, 2006). Au sein du groupe de travail sur les troubles de l'humeur, responsable de la création du TDPM, 75 % des membres avaient des liens avec l'industrie pharmaceutique. Pour justifier la création du TDPM, on peut lire l'argumentaire suivant :

The inclusion of PMDD as a diagnostic category may further facilitate development of treatments that are useful for PMDD and may encourage research into the biology, prevalence, as well as consequences of PMDD [...] It should also be mentioned that there is already some acceptance for PMDD as

an independent category from Federal regulators in that several medications have received an indication for treatment of PMDD. (dsm5.org) (cité dans Cosgrove et Wheeler, 2013b).

Avec cet exemple, on assiste alors à quelque chose comme un « narratif du médicament » (*drug narrative*, voir Matheson, 2008), où la découverte d'un médicament apparaît comme la preuve de la validité d'une catégorie diagnostique. L'influence de l'industrie pharmaceutique sur la médicalisation peut prendre au moins trois formes. L'industrie pharmaceutique peut chercher :

- 1) À assouplir les frontières d'une catégorie déjà existante (Matheson, 2008), comme le trouble de déficit d'attention avec hyperactivité chez les adultes (Conrad, 2007; Conrad et Slodden, 2013);
- 2) À conceptualiser des diagnostics sous la forme de continuum (Matheson, 2008);
- 3) À pathologiser des conditions considérées comme des problèmes normaux de vie (Conrad et Slodden, 2013), par exemple la détresse ressentie par le deuil (Dorwick et Frances, 2013; Horwick et Wakefield, 2007), le TDPM et des formes « douces » de phobie sociale, auparavant considérée comme de la timidité (Lane, 2007).

Ces différentes formes de médicalisation peuvent entraîner des conséquences problématiques comme des effets secondaires non nécessaires, une augmentation du stigma, et une mauvaise allocation des ressources en santé mentale (Bastra et Frances, 2011).

Il apparaît donc clair que l'influence de l'industrie pharmaceutique sur la psychiatrie est un facteur majeur qui pousse à la médicalisation croissante. Le modèle fondé sur l'OS que je propose d'adopter est-il en mesure de faire face à cette influence? Dans un

premier temps, il faut noter que l'OS* ne semble pas recommander l'exclusion de l'industrie pharmaceutique du panel. Étant donné que l'industrie pharmaceutique ne promeut pas de valeurs discriminatoires (allant à l'encontre de la norme d'égalité d'autorité intellectuelle tempérée), et que le point de vue de la recherche en psychopharmacologie est pertinent pour penser la définition du trouble mental, il n'y a pas de bonnes raisons, dans le cadre de l'OS*, d'exclure cette perspective du panel. La perspective d'un représentant possiblement biaisé en faveur de l'industrie pharmaceutique n'est pas incompatible avec l'OS*. Rappelons que dans le cadre de l'OS*, le fait que les membres de la communauté épistémique endossent certaines valeurs qui peuvent influencer leur position n'est pas en soi un problème. Il est donc possible d'imaginer qu'un représentant de la recherche en psychopharmacologie siège sur ce panel, participant ainsi à maximiser la diversité de *standpoints* du panel.

Or l'OS* inclut des ressources pour faire face à l'influence indue de l'industrie pharmaceutique au sein de ce panel, car elle s'inscrit dans une approche de « gestion des valeurs ». En effet, elle vise à identifier et à critiquer les préjugés *personnels* en utilisant le niveau de la *communauté épistémique*. Comme les individus sont généralement mauvais pour reconnaître leurs propres préjugés (puisqu'ils sont souvent implicites), la solution proposée par Longino réside dans la conception de la communauté épistémique : si la communauté est diversifiée de manière adéquate, la multiplicité de perspectives augmente les chances d'identifier et de critiquer les valeurs idiosyncrasiques (car les individus sont plus aptes à détecter les préjugés et les valeurs lorsqu'ils ne sont pas les leurs). Dans le contexte du panel pour définir le trouble mental, c'est la diversité du panel qui est censée bloquer les effets des intérêts commerciaux contestables portés par certains individus (parfois de manière inconsciente). Qui plus est, il faut noter qu'à lui seul, un représentant de l'industrie pharmaceutique ne pourrait pas avoir le contrôle entier sur la définition du trouble mental, puisque les prises de décisions au sein du panel se feraient par un vote. Bien que Browne rejette le consensus parce qu'elle le trouve peu réaliste sur les questions

de valeurs, le cas de l'industrie pharmaceutique me semble constituer un argument supplémentaire en faveur du vote, et contre le consensus. En effet, grâce à la méthode du vote, il n'est pas possible pour le représentant d'une perspective possiblement problématique de bloquer à lui seul une décision. En outre, Cosgrove et Wheeler (2013a) ont défendu qu'une vision de la science indépendante des valeurs viendrait masquer l'impact des intérêts motivés par le profit sur la science. Puisque l'OS* se fonde sur le rejet de cette vision de la science indépendante des valeurs et plaide plutôt pour la gestion des valeurs, elle représente une solution réaliste aux problèmes engendrés par les relations financières de la psychiatrie avec l'industrie pharmaceutique. Avec le panel, il devient ainsi possible de mettre en évidence et de contrôler des intérêts motivés par le profit, qui ne devraient pas entrer dans le processus.

À ce stade, il convient toutefois de noter que le panel envisagé ne donne pas de garantie de succès pour empêcher toute forme d'influence indue motivée par le profit sur la définition du trouble mental. La présence d'importants conflits d'intérêts chez les spécialistes de la santé mentale pourrait venir miner la force de la diversité, caractéristique de l'OS*. En effet, si plusieurs représentants siégeant sur le panel ont des liens avec l'industrie pharmaceutique (comme c'est le cas actuellement dans les groupes de travail de la révision du DSM), l'OS* devient moins efficace. Pour l'instant, l'APA a mis en place deux mesures pour contrer l'omniprésence des conflits d'intérêts : la divulgation des conflits d'intérêts financiers au cours des trois dernières années et les restrictions des liens actuels avec l'industrie (pour une discussion, voir Cosgrove *et al.*, 2014; PLoS Medicine Editors, 2012). Ces mesures semblent cependant insuffisantes pour assurer une réelle diversité au sein du panel. Il est donc possible de demander que le nombre de représentants ayant des liens avec l'industrie pharmaceutique soit limité, voire de demander qu'outre le représentant de l'industrie pharmaceutique, les autres membres du panel ne soient pas en situation de conflit d'intérêts. Ainsi, le cadre de l'OS* donne certains moyens pour lutter contre l'influence indue de l'industrie pharmaceutique sur la définition du trouble mental. Or comme je

l'ai déjà laissé sous-entendre, l'influence de l'industrie pharmaceutique sur la psychiatrie est un problème important, et il va de soi que ce seul panel ne constitue pas à lui seul un remède à cette influence.

Conclusion

Dans ce chapitre, j'ai montré que bien que les processus de révision généraux du DSM-5 étaient plus inclusifs et transparents qu'auparavant, le processus de révision de la définition du trouble mental présentait plusieurs limites. Une limite importante est l'absence de mécanismes structurels pour discuter les enjeux de valeurs liés à la définition. Avec le cadre de l'OS*, j'ai pu suggérer une réforme dans le but de pallier ces problèmes : un panel diversifié responsable des enjeux de valeurs, dont la définition du trouble mental. Ce panel serait en mesure d'assurer un dialogue critique quant aux valeurs qui peuvent influencer légitimement la définition. La création de ce panel permettrait au DSM de prétendre à un concept de trouble mental qui serait à la fois socialement construit et objectif.

Il est cependant possible d'objecter que la manière d'appliquer ce panel aux structures de révisions du DSM demeure floue. En effet, certains paramètres de ce panel sont toujours à débattre, comme la manière de choisir les membres du panel, sa composition exacte, son mode de délibération, la manière de faire face à des désaccords importants, etc. Il s'agit là, en mon sens, d'une limite propre aux théories de l'OS. Comme je l'ai mentionné en début de chapitre, l'OS* est une théorie très générale sur la manière de faire face aux valeurs en science. Or, elle n'est pas suffisante pour guider une

application précise dans un contexte particulier d'investigation scientifique. Des ressources peuvent cependant être trouvées en dehors du cadre de l'OS.

À titre d'exemple, les écrits sur l'éthique de la discussion (p. ex. : Habermas, 1996; Rawls, 1971) et les processus démocratiques de délibération en science me semblent particulièrement intéressants (voir Bächtiger *et al.*, 2018; Bohman, 2006; Martí, 2006; voir Karin, 2010 pour une application aux sciences environnementales). La démocratie délibérative implique une discussion entre participants.es ayant pour but l'évaluation des valeurs et intérêts liés aux questions d'intérêt commun. Le rôle de la délibération au sein des communautés et des institutions scientifiques est devenu récemment un objet d'étude au sein de ces écrits, notamment en ce qui concerne la production de jugements et de décisions scientifiques impliquant des jugements de valeur. Plusieurs types de mécanismes délibératifs ont été discutés (p. ex. : les panels de citoyens, les conférences de consensus et les « mini-publics », soit un groupe restreint de citoyen choisi de manière aléatoire devant élaborer des recommandations « représentatives » de la population en général, voir Moore, 2018 pour une discussion de ces différents mécanismes). Or pour l'instant, ces écrits n'ont pas été liés aux théories de l'objectivité sociale en science (une exception est peut-être Solomon, 2015, mais ce travail reste très exploratoire). Analyser chacun de ces mécanismes et leur potentiel d'application au processus de révision du DSM conformément à l'idéal de l'OS* permettrait de contribuer à lier ces deux types d'écrits et à rendre l'OS* plus concrète.

Lier l'éthique et l'épistémologie de la représentation (p. ex. : Bohman, 2012; Brito Vieira et Runciman, 2008; Brown, 2009) aux théories de l'objectivité sociale apparaît également comme une piste fructueuse pour penser l'applicabilité de l'OS*. L'OS* implique que des participants.es représentatifs.ves de certains *standpoints* participent à la discussion critique. Déterminer comment les participants.es peuvent s'exprimer au nom de leur *standpoint* apparaît comme un enjeu central, d'autant plus que ce type de problème a déjà surgi en psychiatrie. En effet, certains représentants des groupes de

patients sont récemment devenus la cible de critique, leurs opposants qualifiant cette représentation de « partielle » (voir Chapman et Veit, 2020; McCoy *et al.*, 2020 au sujet de la représentation du spectre de l'autisme par des groupes de défense des patients).

Qui plus est, il y a fort à parier que des divergences vont émerger quant à la « bonne » définition du trouble mental en fonction des différentes positions informées par les *standpoints* des participants.es (p. ex. : les psychiatres pourraient avoir davantage tendance à adopter une définition plus inclusive du trouble, dans le but de minimiser les faux négatifs, alors que des approches en sciences humaines pourraient avoir tendance à préférer une définition plus étroite, pour pouvoir lutter contre la surmédicalisation). Le lien entre pluralisme et objectivité devrait donc également être exploré plus avant (Longino, 2002, chapitre 8; pour une analyse des enjeux liés au pluralisme en psychiatrie, voir Bueter, 2019 et Tsou, 2015; voir Solomon, 2015 pour un argument contre le pluralisme dans la recherche appliquée).

Cela étant dit, malgré l'incomplétude de la théorie de l'OS quant à son applicabilité, il est juste de dire que le panel que je propose constitue une première ébauche de ce à quoi pourrait ressembler un DSM socialement objectif au regard de sa définition du trouble mental. Comme je l'ai défendu, ce panel a le potentiel de rendre explicites, discutables et critiquables les valeurs en jeu dans la définition, ce qui manque actuellement au processus de révision du DSM. Cependant, préciser les paramètres exacts du panel demandera de sortir du cadre de l'OS et d'emprunter à d'autres approches, comme celles évoquées plus haut. Ce sera l'objet de recherches futures.

Références du chapitre 7

- Alarcón, R. D. (2014). Cultural inroads in DSM-5. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, 13(3), 310-313.
- Amoretti, M. C. et Lalumera, E. (2019). A potential tension in DSM-5 : The general definition of mental disorder versus some specific diagnostic criteria. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 44(1), 85-108.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. (5^e édition). Arlington, VA.
- Bächtiger, A., Dryzek, J. S., Mansbridge, J. et Warren, M. E. (dir.). (2018). *The Oxford handbook of deliberative democracy*. Oxford : Oxford University Press.
- Batstra, L. et Frances, A. (2012). Holding the line against diagnostic inflation in psychiatry. *Psychotherapy and Psychosomatics*, 81(1), 5-10.
- Bell, S. E. et Figert, A. E. (2012). Medicalization and pharmaceuticalization at the intersections : Looking backward, sideways and forward. *Social Science & Medicine*, 75(5), 775-783.
- Biddle, J. (2007). Lessons from the Vioxx debacle : What the privatization of science can teach us about social epistemology. *Social Epistemology*, 21(1), 21-39.
- Blease, C. (2017). Philosophy's territorialism : Scientists can talk about values too. *Philosophy, Psychiatry, & Psychology*, 24(3), 231-234.
- Bohman, J. (2006). Deliberative democracy and the epistemic benefits of diversity. *Episteme : A Journal of Social Epistemology*, 3(2), 175-191.

- Bohman, J. (2012). Representation in the deliberative system. Dans J. Parkinson et J. Mansbridge (dir.), *Deliberative systems : Deliberative democracy at the large scale* (p. 72-94). Cambridge : Cambridge University Press.
- Brito Vieira, M. et Runciman, D. (2008). *Representation*. Cambridge : Polity Press.
- Brown, M. B. (2009). *Science in democracy : Expertise, institutions and representation*. Cambridge, MA : MIT Press.
- Browne, T. K. (2017a). A role for philosophers, sociologists and bioethicists in revising the DSM : A philosophical case conference. *Philosophy, Psychiatry, & Psychology*, 24(3), 187-201.
- Browne, T. K. (2017b). An ethics review panel for the DSM : A worthwhile challenge. *Philosophy, Psychiatry, & Psychology*, 24(3), 235-240.
- Bueter, A. (2018). Public epistemic trustworthiness and the integration of patients in psychiatric classification. *Synthese*. doi.org/10.1007/s11229-018-01913-z
- Bueter, A. (2019). Epistemic injustice and psychiatric classification. *Philosophy of Science*, 86(5), 1064-1074.
- Callaghan, G. M., Chacon, C., Coles, C., Botts, J. et Laraway S. (2009). An empirical evaluation of the diagnostic criteria for premenstrual dysphoric disorder : Problems with sex specificity and validity. *Women & Therapy*, 32(1), 1-21.
- Caplan, P. J., McCurdy-Myers, J. et Gans, M. (1992). Should premenstrual syndrome be called a psychiatric abnormality? *Feminism & Psychology*, 2(1), 27-44.
- Caplan, P. J. (1995). *They say you're crazy : How the world's most powerful psychiatrists decide who's normal*. Reading, MA : Addison-Wesley.
- Caplan, P. J. (2004). The debate about PMDD and Sarafem. *Women & Therapy*, 27(3-4), 55-67.
- Chapman, R. et Veit, W. (2020). Representing the autism spectrum. *The American Journal of Bioethics*, 20(4), 46-48.

- Conrad, P. (2007). *The medicalization of society : On the transformation of human conditions into treatable disorders*. Baltimore : Johns Hopkins University Press.
- Conrad, P. et Slodden, C. (2013). The medicalization of mental disorder. Dans C. S. Aneshensel, J. C. Phelan et A. Bierman (dir.), *Handbook of the sociology of mental health* (p. 61-73). Dordrecht : Springer.
- Cooper, R. (2015). Must disorders cause harm? The changing stance of the DSM. Dans S. Demazeux et P. Singy (dir.), *The DSM-5 in perspective* (p. 83-96). Dordrecht : Springer.
- Cooper, R. (2017). A modest proposal. *Philosophy, Psychiatry, & Psychology*, 24(3), 207-209.
- Cosgrove, L. et Caplan, P. J. (2004). Medicalizing menstrual distress. Dans P. J. Caplan et L. Cosgrove (dir.), *Bias in psychiatric diagnosis* (p. 221-230). Lanham, MD : Jason Aronson.
- Cosgrove, L. et Krinsky, S. (2012). A comparison of DSM-IV and DSM-5 panel members' financial associations with industry : A pernicious problem persists. *PLOS Medecine*, 9(3). doi.org/10.1371/journal.pmed.1001190
- Cosgrove, L. et Riddle, B. (2004). Gender bias and sex distribution of mental disorders in the DSM-IV-TR. Dans P. J. Caplan et L. Cosgrove (dir.), *Bias in psychiatric diagnosis* (p. 127-140). Lanham, MD : Jason Aronson.
- Cosgrove, L. et Wheeler, E. (2013a). Industry's colonization of psychiatry : Ethical and practical implications of financial conflicts of interest in the DSM-5. *Feminism & Psychology*, 23, 93-106.
- Cosgrove, L. et Wheeler, E. E. (2013b). Drug firms, the codification of diagnostic categories, and bias in 929 clinical guidelines. *The Journal of Law, Medicine & Ethics*, 41(3), 644-653.

- Cosgrove, L., Krimsky, S., Vijayaraghavan, M. et Schneider, L. (2006). Financial ties between DSM-IV panel members and the pharmaceutical industry. *Psychotherapy and Psychosomatics*, 75(3), 154-160.
- Cosgrove, L., Krimsky, S., Wheeler, E. E., Kaitz, J., Greenspan, S. B. et DiPentima, N. L. (2014). Tripartite conflicts of interest and high stakes patent extensions in the DSM-5. *Psychotherapy and Psychosomatics*, 83(2), 106-113.
- Di Giulio, G. G. et Reissing, E. D. (2006). Premenstrual Dysphoric Disorder : Prevalence, Diagnostic Considerations, and Controversies. *Journal of Psychosomatic Obstetrics & Gynecology*, 27(4), 201-210.
- Douglas, H. (2003). The moral responsibilities of scientists (tensions between autonomy and responsibility). *American Philosophical Quarterly*, 40(1), 59-68.
- Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh : University of Pittsburgh Press.
- Dowrick, C. et Frances, A. (2013). Medicalising unhappiness : New classification of depression risks more patients being put on drug treatment from which they will not benefit. *British Medical Journal*.
doi.org/10.1371/journal.pmed.1001190
- First, M. B. et Wakefield, J. C. (2013). Diagnostic criteria as dysfunction indicators : Bridging the chasm between the definition of mental disorder and diagnostic criteria for specific disorders. *The Canadian Journal of Psychiatry*, 58(12), 663-669.
- Frances, A. J. et Widiger, T. (2012). Psychiatric diagnosis : Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, 8, 109-130.
- Gureje, O. et Stein, D. J. (2012). Classification of mental disorders : The importance of inclusive decision-making. *International Review of Psychiatry*, 24(6), 606-612.

- Habermas, J. (1992/1996). *Between facts and norms : Contributions to a discourse theory of law and democracy*. (W. Rehg, trad.). Cambridge, MA : MIT Press.
- Haque, O. S., Lu, A., Wu, D., Cosgrove, L. et Bursztajn, H. J. (2014). Curing financial conflicts of interest in psychiatric professional organizations. Dans J. Z. Sadler, B. Fulford et C. W. van Staden (dir.), *The Oxford handbook of psychiatric ethics*. Oxford : Oxford University Press.
- Horwitz, A. V. et Wakefield, J. (2007). *The loss of sadness : How psychiatry transformed normal sorrow into depressive disorder*. New York : Oxford University Press.
- Karin, B. (dir.) (2010). *Environmental politics and deliberative democracy : Examining the promise of new modes of governance*. Northampton, MA : Edward Elgar Publishing.
- Kellert, S., Longino, H. et Waters, K. (2006). Introduction : The pluralist stance. Dans S. Kellert, H. Longino et K. Waters (dir.), *Scientific pluralism* (p. vii–xxix). Minneapolis : University of Minnesota Press.
- Kendler, K. S., Appelbaum, P. S., Bell, C. C., Fulford, K. W. M., Ghaemi, S. N., Schaffner, K. F., ... Sadler, J. Z. (2008). Issues for DSM-V : DSM-V should include a conceptual issues work group. *American Journal of Psychiatry*, 165(2), 174-175.
- Kendler, K. S. et Solomon, M. (2016). Expert consensus v. evidence-based approaches in the revision of the DSM. *Psychological Medicine*, 46(11), 2255-2262.
- Kinghorn, W. (2013). The biopolitics of defining “mental disorder”. Dans J. Parris et J. Phillips (dir.), *Making the DSM-5* (p. 47-61). New York : Springer.
- Kupfer, David J., First, Michael B. et Regier, Darrel A. (dir.). (2002). *A research agenda for DSM-5*. Washington : American Psychiatric Association.
- Lane, C. (2007). *Shyness : How normal behavior became a sickness*. New Haven, CT : Yale University Press.

- Lexchin, J. et O'Donovan, O. (2010). Prohibiting or “managing” conflict of interest? A review of policies and procedures in three European drug regulation agencies. *Social Science and Medicine*, 70, 643-647.
- Martí, J. L. (2006). The epistemic conception of deliberative democracy defended. Dans S. Besson et J. L. Martí (dir.), *Deliberative democracy and its discontents* (p. 27-56). Aldershot : Ashgate.
- Matheson, A. (2008). Corporate science and the husbandry of scientific and medical knowledge by the pharmaceutical industry. *BioSocieties*, 3(4), 355-382.
- McCoy, M. S., Liu, E. Y., Lutz, A. S. et Sisti, D. (2020). Ethical advocacy across the autism spectrum : Beyond partial representation. *The American Journal of Bioethics*, 20(4), 13-24.
- Moon, M. R. et Khin-Maung-Gyi, F. (2009). The history and role of institutional review boards. *Virtual Mentor*, 11(4), 311-316.
- Moynihan, R. N., Cooke, G. P., Doust, J. A., Bero, L., Hill, S. et Glasziou, P. P. (2013). Expanding disease definitions in guidelines and expert panel ties to industry : A cross-sectional study of common conditions in the United States. *PLOS Medicine*, 10(8).
- Musschenga, A. W., Van der Steen, Wim J. et Ho, Vincent K. Y. (2010). The Business of Drug Research : A Mixed Blessing. Dans H. Radder (dir.), *The commodification of academic science* (p. 110-131). Pittsburgh : Pittsburgh University Press.
- Narrow, W. E. et Kuhl, E. A. (2010). Clinical significance and disorder thresholds in DSM-5. Dans D. A. Regier, W. E. Narrow, E. A. Kuhl et D. J. Kupfe (dir.), *The conceptual evolution of DSM-5* (p. 147-162). Arlington, VA: American Psychiatric Publishing.
- Offman, A. et Kleinplatz, P. J. (2004). Does PMDD belong in the DSM? Challenging the medicalization of women's bodies. *Canadian Journal of Human Sexuality*, 13(1), 17-27.

- Phillips, J., Frances, A., Cerullo, M. A., Chardavoyne, J., Decker, H. S., First, M. B., ... Zachar, P. (2012). The six most essential questions in psychiatric diagnosis : A pluralogue. Part 1 : Conceptual and definitional issues in psychiatric diagnosis. *Philosophy, Ethics and Humanities in Medicine*, 7(3), 1-29.
- Pilecki, B. C., Clegg, J. W. et McKay, D. (2011). The influence of corporate and political interests on models of illness in the evolution of the DSM. *European Psychiatry*, 26(3), 194-200.
- PLOS Medicine Editors (2012). Does conflict of interest disclosure worsen bias? *PLOS Medicine*, 9(4). doi.org/10.1371/journal.pmed.1001210
- Porter, D. (2017). Ontological assumptions, a biopsychosocial approach, and patient participation : Moving toward an ethically legitimate science of psychiatric nosology. *Philosophy, Psychiatry, & Psychology*, 24(3), 223-226.
- Pouncey, C. et Merz, J. F. (2017). Browne's external DSM ethical review panel : That dog won't hunt. *Philosophy, Psychiatry, & Psychology*, 24(3), 227-230.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA : Harvard University Press.
- Regier, D. A., Kuhl, E. A., Kupfer, D. J. et McNulty, J. P. (2010). Patient involvement in the development of DSM-V. *Psychiatry : Interpersonal and Biological Processes*, 73(4), 308-310.
- Regier, D. A., Narrow, W. E., First, M. B. et Marshall, T. (2002). The APA classification of mental disorders : Future perspectives. *Psychopathology*, 35(2-3), 166-170.
- Regier, D. A., Narrow, W. E., Kuhl, E. A., Kupfer, D. J. et American Psychopathological Association (2011). *The conceptual evolution of DSM-5*. Arlington, VA : American Psychiatric Publishing.

- Ross, P. A. (2002). Values and objectivity in psychiatric nosology. Dans J. Z. Sadler (dir.), *Descriptions and prescriptions : Values, mental disorders, and the DSMs* (p. 45-55). Baltimore : John Hopkins University Press.
- Rounsaville, B. J., Alarcón, R. D., Andrews, G., Jackson, J. S., Kendell, R. E. et Kendler, K. (2002). Basic nomenclature issues for DSM-V. Dans D. J. Kupfer, M. B. First et D. S. Regier (dir.), *A research agenda for DSM-V* (p. 1-30). Washington : American Psychiatric Association.
- Sadler, J. Z. (2002). Values in developing psychiatric classifications : A proposal for DSM-V. Dans J. Z. Sadler (dir.), *Descriptions and prescriptions : Values, mental disorders, and the DSMs* (p. 301-322). Baltimore : John Hopkins University Press.
- Sadler, J. Z. et Fulford, B. (2004). Should patients and their families contribute to the DSM-V process? *Psychiatric Services*, 55(2), 133-138.
- Sisti, D. et Johnson, R. (2015). Revision and representation : The controversial case of DSM-5. *Public Affairs Quarterly*, 29(1), 76-108.
- Sobel, M. et Hansen, K. (2006). Banking of human biological materials for research. Dans E. Bankert et R. Amdur (dir.), *Institutional review board : Management and function* (p. 433-440). Sudbury, MA : John and Bartlett.
- Spitzer, R. L. (2009). DSM-V transparency : Fact or rhetoric? *Psychiatric Times*, 26(3).
- Spitzer, R. L. et Wakefield, J. C. (1999). DSM-IV diagnostic criterion for clinical significance : Does it help solve the false positives problem? *American Journal of Psychiatry*, 156(12), 1856-1864.
- Stein, D. J. (2014). Psychopharmacology and natural kinds : A conceptual framework. Dans S. Kincaid et K. Sullivan (dir.), *Classifying psychopathology : Mental kinds and natural kinds* (p. 65-73). Cambridge, MA : MIT Press.

- Stein, D. J. et Phillips, K. A. (2013). Patient advocacy and DSM-5. *BMC Medicine*, 11(1), 133-138.
- Tseris, E. (2018). Biomedicine, neoliberalism and pharmaceuticalisation of society. Dans A. V. Horwitz et A. Scull, *The Routledge international handbook of critical mental health* (p. 169-176). New York : Routledge.
- Williams, S. J., Martin, P. et Gabe, J. (2011). The pharmaceuticalisation of society? A framework for analysis. *Sociology of Health and Illness*, 33(5), 710-725.
- Yager, J. et McIntyre, J. S. (2014). DSM-5 clinical and public health committee : Challenges and considerations. *American Journal of Psychiatry*, 171(2), 142-144.

CONCLUSION GÉNÉRALE

UN NORMATIVISME OBJECTIF POUR DÉFINIR LE TROUBLE MENTAL

Le point de départ de cette thèse était l'émergence de plusieurs litiges ayant impliqué la psychiatrie américaine au fil de son histoire : les attaques du mouvement antipsychiatrique, les catégories diagnostiques controversées (comme l'homosexualité ou le trouble dysphorique prémenstruel), les critiques de surmédicalisation et le mouvement de la neurodiversité. Toutes soulevaient l'enjeu de la définition du trouble mental ainsi que le rôle des valeurs dans cette définition. Une stratégie importante de la réponse à cet enjeu, du point de vue philosophique du moins, a été d'invoquer la présence d'une norme biologique. Si les fonctions naturelles de l'esprit humain peuvent être découvertes uniquement par l'observation empirique, il serait possible de se positionner vis-à-vis cet enjeu sans faire appel à des valeurs, possiblement instables ou impliquant d'importants désaccords. Qui plus est, faire appel à une notion indépendante de celle des valeurs viendrait consolider le statut scientifique de la psychiatrie, lui aussi malmené par plusieurs de ces attaques.

Le premier objectif de cette thèse était de démontrer qu'il est peu plausible de parvenir à démarquer le fonctionnement mental normal du pathologique grâce à une norme physiologique. Une composante purement indépendante des valeurs, généralement celle de dysfonction, n'est pas capable de définir le trouble mental de manière satisfaisante. La première partie de cette thèse a donc été consacrée à l'examen critique de l'approche objectiviste des fonctions naturelles, qui défend que la notion de

dysfonction est ce qui assure un fondement scientifique et objectif au concept de trouble mental. J'ai montré que l'abondance et la justesse des critiques ayant visé cette approche, sous la forme de la BST ou de la DP, devraient conduire à son abandon. Dans leur état actuel, il semble que ces approches ne puissent pas jouer le rôle salvateur qu'on leur avait initialement attribué, celui d'assurer l'objectivité et la neutralité du concept. La réfutation de l'approche objectiviste des fonctions naturelles mène à la première conclusion de cette thèse : les valeurs jouent un rôle majeur dans le processus de démarcation entre les états mentaux sains et ceux pathologiques.

Le second objectif de cette thèse était de montrer que le cadre de l'OS* permet de rendre compte de l'interaction entre faits et valeurs caractéristique du concept de trouble mental utilisé en psychiatrie américaine. En suivant le cadre de l'OS*, l'objectivité fait référence aux processus par lesquels la définition est produite (objectivité sociale) ainsi qu'au contenu empirique associé à la définition (objectivité empirique). L'adoption de l'OS* est basée sur le rejet de l'une des présuppositions centrales de l'approche objectiviste des fonctions naturelles, soit l'idéal de la science indépendante des valeurs. Une fois cet idéal rejeté, le nouvel idéal de l'OS* apparaît comme réaliste et profitable pour penser la question des valeurs qui influencent le concept de trouble mental. Selon l'OS*, l'influence des valeurs sur le concept de trouble mental n'est pas un problème en soi. Le problème réside plutôt dans le fait que ces valeurs ne sont, à l'heure actuelle, ni explicitées, ni examinées de manière critique. La défense de l'utilisation de l'OS* pour penser la définition du trouble mental mène à la seconde conclusion de cette thèse : la définition du trouble mental peut être dite à la fois socialement construite et objective si les standards de l'OS* sont respectés.

Adopter l'OS* pour penser cette définition permet de mieux faire face aux controverses liées à la question de la définition du trouble mental auxquelles est confrontée la psychiatrie contemporaine depuis les années 1950. Dans un premier temps, l'OS* est un meilleur cadre pour faire face à des controverses en ce qu'elles ont de social, c'est-

à-dire en ce qu'elles impliquent divers acteurs provenant de situations sociales différentes. Le rapport de force du corps médical vis-à-vis les patients.es, mais aussi la subordination de certains groupes sociaux comme les femmes et les personnes racisées vis-à-vis l'autorité psychiatrique peut être prise en compte grâce à la théorie de l'OS*. Resituer le débat sur la définition du trouble mental au niveau des structures et des acteurs qui l'ont fait émerger ou qui devraient l'élaborer permet de mieux penser la place et le rôle des valeurs dans cette question. En effet, cette approche tient pour acquise la présence probable de valeurs en science, admet la possibilité de perspectives divergentes sur un sujet donné et donne un cadre pour gérer la coexistence de plusieurs visions et attitudes divergentes.

Dans un second temps, l'OS* permet de répondre au problème de la scientificité de la psychiatrie lié à la définition du trouble mental. L'OS* rend compte des multiples degrés d'objectivité associés à différentes sciences (p. ex. : la psychiatrie par contraste avec la chimie organique). Les attaques du mouvement de l'antipsychiatrie avaient remis en question la légitimité scientifique de la psychiatrie. La psychiatrie et certains courants de la philosophie de la psychiatrie ont souhaité, depuis les années 1970, se rapprocher d'une objectivité « dure », au sens de l'indépendance des valeurs, pensant que c'était la seule manière de se tirer des problèmes qu'elle rencontrait. La présence de valeurs au sein du concept de trouble mental avait donc été perçue avec une grande suspicion par plusieurs. Le cadre de l'OS* permet de montrer comment on peut résoudre certains problèmes liés au caractère normatif de la psychiatrie et de l'un de ses concepts centraux, le trouble mental, sans pour autant devoir la rapprocher d'une science parfaitement indépendante des valeurs. Bien sûr, dans le futur, des approches naturalistes du trouble mental pourraient être développées, et des arguments philosophiques convaincants avancés pour démontrer la possible neutralité du trouble mental, ou du moins de la notion de dysfonction. Ce que cette thèse montre, c'est que les efforts de ces approches ne sont pas nécessaires pour assurer à la psychiatrie cette légitimité scientifique. La quête de l'indépendance des valeurs d'une des composantes

de la définition du trouble mental n'est pas aussi centrale qu'il y paraît pour sauver l'objectivité et la scientificité de la psychiatrie.

Du point de vue de ses implications, l'une des conséquences majeures de cette thèse est que si la proposition d'intégrer le cadre de l'OS* est jugée convaincante, le débat traditionnel sur la définition du trouble mental est amené à se métamorphoser. En effet, l'enjeu du *contenu* de la définition devient moins important que celui du *processus* par lequel cette définition devrait émerger. En effet, en adoptant le cadre de l'OS*, évaluer une définition du trouble mental équivaut à juger la manière dont la définition a émergé et si elle respecte les standards de l'OS*. Il s'agit moins de savoir si le trouble mental devrait se définir par la notion de dysfonction, de détresse ou de trouble de l'agentivité que de déterminer si des structures permettant le dialogue critique et diversifié ont été mises en place pour que la définition soit élaborée. L'OS* attire l'attention sur les structures et les processus sociaux qui permettent la production de connaissances scientifiques, plutôt que sur son contenu. Suivant ce nouveau cadre, le concept de trouble mental pourrait être objectif si la manière dont la définition est élaborée permet l'intégration de la critique des différents groupes touchés par cette définition (groupes de patients et de familles, chercheurs en sciences humaines, psychiatres et autres professionnels de la santé mentale, décideurs publics, etc.). Mettre en relation cette nouvelle approche de l'objectivité sociale avec les écrits sur le concept de trouble mental permet de sortir la philosophie de la psychiatrie d'un débat de longue date quant à la scientificité du concept de trouble mental, débat qui offrait jusqu'alors peu d'espoir de résolution.

Cependant, s'il faut revenir à la catégorisation traditionnelle de ce débat, on peut se demander où situer cette thèse au regard de l'opposition entre l'approche objectiviste des fonctions naturelles et le normativisme. La position soutenue dans cette thèse est une forme de compromis entre ces deux pôles. Elle est normativiste parce qu'ayant rejeté que la notion de dysfonction pouvait fonder une distinction naturelle entre le

normal et le pathologique, ce sont les valeurs qui viennent entièrement assumer cette tâche. Or il ne s'agit pas d'un normativisme traditionnel, au sens où l'argument fourni dans cette thèse permet de *légitimer* la présence de valeurs sociales et culturelles dans la définition sans pour autant être conduit à abandonner le statut scientifique de la psychiatrie. Cette légitimation était absente des approches normativistes classiques. Le normativisme devient une position tenable si l'on veut penser le concept de trouble mental sur le plan scientifique, moyennant l'adoption de l'OS*. Cette thèse est donc aussi une tentative de dépassement du normativisme traditionnel : dans le cadre de l'OS*, les valeurs peuvent légitimement influencer le concept de trouble mental si elles sont discutées et critiquées par un ensemble de *standpoints* différents.

Mais la position défendue dans cette thèse, bien qu'avant tout normativiste, n'équivaut pas à un rejet total de l'approche objectiviste des fonctions naturelles. Elle en conserve deux intuitions. D'abord, qu'il est essentiel pour la psychiatrie de fonder la distinction entre le normal et le pathologique sur un ou des critères objectifs si elle veut répondre à ses critiques. Ensuite, que la notion de dysfonction, parce qu'elle semble jouer un rôle important dans le discours théorique médical, peut conserver sa place. Néanmoins, cette idée de dysfonction est repensée dans le cadre de l'OS*. Elle est ce qui assure le contenu *empirique* d'une définition du trouble mental, par l'observation de mécanismes causaux expliquant des dysfonctions particulières. La dysfonction n'est plus ce qui vient démarquer le normal du pathologique, tâche occupée par des valeurs discutées et critiquées grâce à la mise en place de structures conformes à l'OS*.

Alors, l'approche défendue dans cette thèse est-elle objectiviste ou normativiste? Elle se situe entre les deux. Elle conserve la thèse normativiste, selon laquelle le critère de démarcation entre le normal et le pathologique provient de normes sociales et culturelles, mais endosse les thèses objectivistes selon lesquelles la notion de dysfonction et celle de trouble mental sont liées, ainsi que l'idéal d'objectivité du concept. Cependant, malgré le fait que l'utilisation de l'OS* dans ce débat rejoigne

plusieurs intuitions initiales de ces approches, elle implique une divergence majeure vis-à-vis la manière de répondre à la question de la définition du trouble mental et la place de la philosophie de la psychiatrie dans ce débat. Comme les lecteurs auront pu le constater, cette thèse ne propose pas de définition alternative du trouble mental. On ne trouve pas de défense de quelconque critère définitionnel comme la détresse ou la dysfonction (bien que j'aie défendu au chapitre 6 que la notion de dysfonction pourrait trouver sa place dans une définition du trouble mental, si réinterprétée dans le cadre de l'OS*). Il faut insister à nouveau sur le fait que l'OS* propose des *moyens structurels* de parvenir à une définition objective du trouble, mais pas de contenu définitionnel. Le rôle des philosophes s'en trouve alors considérablement modifié. L'OS* ne demande pas aux philosophes de réfléchir aux critères définitionnels du trouble mental, car ce rôle est désormais partagé entre plusieurs acteurs touchés par cette définition. La voix des philosophes de la psychiatrie est donc limitée, en ce qu'elle devient une voix parmi d'autres pouvant s'exprimer sur la question. L'OS* place la réflexion philosophique du côté des structures de prise de décisions appropriées pour résoudre la question, et non plus du côté de « la bonne » définition du trouble mental. Avec le débat traditionnel qui se métamorphose, la fonction des philosophes se transforme elle aussi.

Au-delà de cette transformation du débat traditionnel sur la définition du trouble mental, importer le cadre de l'OS* à la définition du trouble mental ouvre de nouvelles questions de recherche pour la philosophie de la psychiatrie. Comme l'analyse des structures de prises de décisions de l'APA présentée au chapitre 7 a permis de le montrer, l'OS* ne donne pas d'indications précises quant à la manière de procéder pour son application. L'OS* ne répond pas à la question « comment mettre en place de manière concrète la délibération critique qui assurerait à la définition du trouble mental son objectivité? ». L'OS* fournit plutôt des arguments convaincants en faveur de la mise en place de ce type de procédés délibératifs en science lorsque des questions de valeurs interagissent avec des faits empiriques. Le choix précis de ces procédés dépasse le cadre de l'OS*. Comme je l'ai laissé sous-entendre à la fin du chapitre 7, en raison

du caractère théorique et abstrait de l'OS*, il convient de se questionner sur la manière dont d'autres champs de recherche tels que la démocratisation en science et l'éthique de la discussion peuvent contribuer à l'applicabilité de l'OS* à la définition du trouble mental. La mise en relation de ces deux champs de recherche avec la philosophie de la psychiatrie paraît prometteuse. Outre l'applicabilité de l'OS* aux structures de révision du DSM, d'autres questions émergent suite à cette thèse. Par exemple, où situer l'OS* vis-à-vis la méthodologie classique de l'analyse conceptuelle employée initialement pour définir le trouble mental? À quel point peut-on être révisionniste vis-à-vis la définition actuelle du trouble mental? Si l'on porte attention aux relations de pouvoir en jeu dans la définition du trouble mental, est-il possible d'identifier des injustices épistémiques dans le processus de définition, c'est-à-dire des injustices qui affectent les individus dans leur manière de connaître et de transmettre la connaissance? Comment penser un pluralisme de la définition du trouble mental? Quelles sont les implications de ce pluralisme pour l'unité de la psychiatrie comme discipline? Lorsqu'appliqué à la psychiatrie et à la définition du trouble mental, le cadre de l'OS* dessine un horizon de recherche nouveau et stimulant. J'espère que cette thèse aura pu contribuer à faire un premier pas dans cette direction.

BIBLIOGRAPHIE

- Agich, G. J. (1983). Disease and value : A rejection of the value-neutrality thesis. *Theoretical Medicine*, 4(1), 27-41.
- Agich, G. J. (1997). Toward a pragmatic theory of disease. Dans J. M. Humber et R. F. Almeder (dir.), *What is disease?* (p. 219-246). Totowa, NJ : Humana Press.
- Agich, G. J. (2002). Implications of a pragmatic theory of disease for the DSMs. Dans J. Z. Sadler (dir.), *Descriptions and prescriptions : Values, mental disorders, and the DSMs* (p. 96-113). Baltimore : Johns Hopkins University Press.
- Alarcón, R. D. (2014). Cultural inroads in DSM-5. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, 13(3), 310-313.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders*. (3rd ed.). Washington : American Psychiatric Association.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders*. (3rd ed., revised). Washington : American Psychiatric Association.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. (4th ed.). Washington : American Psychiatric Association.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. (4th ed., text revision). Washington : American Psychiatric Association.

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. (5th ed.). Washington : American Psychiatric Association.
- Amoretti, M. C. et Vassallo, N. (2015). Against sex and gender dualism in gender-specific medicine. Dans U. Mäki, I. Votsis, S. Ruphy et G. Schurz (dir.), *Recent developments in the philosophy of science : EPSA13 Helsinki* (p. 357-367). s.l. (Suisse) : Springer, Cham.
- Amoretti, M. C. et Vassallo, N. (2017). Introduction : Philosophy of sex and gender in gender medicine. *Topoi*, 36(3), 473-477.
- Amoretti, M. C. et Lalumera, E. (2019). A potential tension in DSM-5 : The general definition of mental disorder versus some specific diagnostic criteria. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 44(1), 85-108.
- Ananth, M. (2008). *In defense of an evolutionary concept of health : Nature, norms, and human biology*. Burlington, VT : Ashgate Publishing Group.
- Anderson, E. (2019). Feminist epistemology and philosophy of science. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/sum2019/entries/feminism-epistemology/>
- Aragona, M. (2009). The concept of mental disorder and the DSM-V. *Dialogues in Philosophy, Mental and Neuro Sciences*, 2(1), 1-14.
- Armstrong, T. (2010). *Neurodiversity : Discovering the extraordinary gifts of autism, ADHD, dyslexia, and other brain differences*. Cambridge, MA : Da Capo Press.
- Aucouturier, V. et Demazeux, S. (2013). The concept of “mental disorder”. Dans H. Carel et R. Cooper (dir.), *Health, illness and disease* (p. 89-106). Londres : Routledge.
- Bächtiger, A., Dryzek, J. S., Mansbridge, J. et Warren, M. E. (dir.). (2018). *The Oxford handbook of deliberative democracy*. Oxford : Oxford University Press.

- Baghramian, M. et Carter, J. A. (2019). Relativism. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/win2019/entries/relativism/>
- Bandura, A., Caprara, G. V., Barbaranelli, C., Gerbino, M. et Pastorelli, C. (2003). Role of affective self-regulatory efficacy in diverse spheres of psychosocial functioning. *Child Development*, 74(3), 769-782.
- Batstra, L. et Frances, A. (2012). Holding the line against diagnostic inflation in psychiatry. *Psychotherapy and Psychosomatics*, 81(1), 5-10.
- Bayer, R. (1981). *Homosexuality and American psychiatry : The politics of diagnosis*. New York : Basic Books.
- Bear, M. F., Connors, B. W. et Paradiso, M. E. (2007). *Neurosciences : À la découverte du cerveau*. (3^e édition). Rueil-Malmaison : Éditions Pradel.
- Bechtel, W. et Abrahamsen, A. (2005). Explanation : A mechanist alternative. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421-441.
- Bechtel, W. et R. C. Richardson (1993/2010). *Discovering complexity : Decomposition and localization as strategies in scientific research*. (2^e édition). Cambridge, MA : MIT Press/Bradford Books.
- Belkhir, J. (1994). Race, sex, class & “intelligence” scientific racism, sexism & classism. *Race, Sex & Class*, 1(2), 53-83.
- Bell, S. E. et Figert, A. E. (2012). Medicalization and pharmaceuticalization at the intersections : Looking backward, sideways and forward. *Social Science & Medicine*, 75(5), 775-783.
- Bergner, R. M. (1997). What is psychopathology? And so what? *Clinical Psychology : Science and Practice*, 4(3), 235-248.

- Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, 3(2), 207-220.
- Biddle, J. (2007). Lessons from the Vioxx debacle : What the privatization of science can teach us about social epistemology. *Social Epistemology*, 21(1), 21-39.
- Biddle, J. (2013). State of the field : Transient underdetermination and values in science. *Studies in History and Philosophy of Science Part A*, 44(1), 124-133.
- Biddle, J. et Kukla, R. (2017). The geography of epistemic risk. Dans K. C. Elliott et T. Richards (dir.), *Exploring inductive risk : Case studies of values in science* (p. 215-238). New York : Oxford University Press.
- Bingham, R. et Banner, N. (2014). The definition of mental disorder : Evolving but dysfunctional? *Journal of Medical Ethics*, 40(8), 537-542.
- Blackless, M. (2000). Recommendations from interACT : Advocates for intersex youth regarding the List of issues for the United States for the 59th session of the Committee against torture. *Human Biology*, 12, 151-166.
- Blackless, M., Charuvastra, A., Derryck, A., Fausto-Sterling, A., Lauzanne, K. et Lee, E. (2000). How sexually dimorphic are we? Review and synthesis. *American Journal of Human Biology*, 12(2), 151-166.
- Blease, C. (2017). Philosophy's territorialism : Scientists can talk about values too. *Philosophy, Psychiatry, & Psychology*, 24(3), 231-234.
- Boghossian, P. (1996, 13 décembre). What the Sokal hoax ought to teach us. *Times Literary Supplement*, 14-15.
- Boghossian, P. (2006). *Fear of knowledge : Against relativism and constructivism*. New York : Clarendon Press.
- Bohman, J. (2006). Deliberative democracy and the epistemic benefits of diversity. *Episteme : A Journal of Social Epistemology*, 3(2), 175-191.

- Bohman, J. (2012). Representation in the deliberative system. Dans J. Parkinson et J. Mansbridge (dir.), *Deliberative systems : Deliberative democracy at the large scale* (p. 72-94). Cambridge : Cambridge University Press.
- Bolton, D. (2001). Problems in the definition of “mental disorder”. *The Philosophical Quarterly*, 51(203), 182-199.
- Bolton, D. (2008). What is mental disorder? An essay in philosophy, science, and values. New York : Oxford University Press.
- Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & Public Affairs*, 5(1), 49-68.
- Boorse, C. (1976a). What a theory of mental health should be. *Journal for the Theory of Social Behaviour*, 6(1), 61-84.
- Boorse, C. (1976b). Wright on functions. *The Philosophical Review*, 85(1), 70-86.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44(4), 542-573.
- Boorse, C. (1987). Concepts of health. Dans D. Van de Veer et T. Regan (dir.), *Health care ethics : An introduction* (p. 359-393). Philadelphie : Temple University Press.
- Boorse, C. (1997). A rebuttal on health. Dans J. M. Humber et R. F. Almeder (dir.), *What is disease?* (p. 1-134). Totowa (New Jersey) : Humana Press.
- Boorse, C. (2002). A rebuttal on functions. Dans A. R. Ariew, R. Cummins et M. Perlman (dir.), *Functions : New essays in the philosophy of psychology and biology* (p. 63-113). Oxford : Oxford University Press.
- Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, 39(6), 683-724.
- Borgerson, K. (2011). Amending and defending critical contextual empiricism. *European Journal for Philosophy of Science*, 1(3), 435-449.

- Boysen, G. A. et Ebersole, A. (2014). Expansion of the concept of mental disorder in the DSM-5. *The Journal of Mind and Behavior*, 35(4), 225-243.
- Boysen, G., Ebersole, A., Casner, R. et Coston, N. (2014). Gendered mental disorders : Masculine and feminine stereotypes about mental disorders and their relation to stigma. *The Journal of Social Psychology*, 154(6), 546-565.
- Bracken, P. et Thomas, P. (2010). From Szasz to Foucault : On the role of critical psychiatry. *Philosophy, Psychiatry, & Psychology*, 17(3), 219-228.
- Briester, E. (2017). Feminism and contextualism. Dans J. J. Ichikawa (dir.), *The Routledge Handbook of Epistemic Contextualism* (p. 57-67). New York : Routledge.
- Briester, E. (2018). Objectivity in science : The impact of feminist accounts. Dans P. Garavaso, *The Bloomsbury companion to analytic feminism* (p. 212-235). Londres : Bloomsbury Academic.
- Brito Vieira, M. et Runciman, D. (2008). *Representation*. Cambridge : Polity Press.
- Brown, M. B. (2009). *Science in democracy : Expertise, institutions and representation*. Cambridge, MA : MIT Press.
- Brown, M. J. (2013). Values in science beyond underdetermination and inductive risk. *Philosophy of Science*, 80(5), 829-839.
- Browne, T. K. (2017a). A role for philosophers, sociologists and bioethicists in revising the DSM : A philosophical case conference. *Philosophy, Psychiatry, & Psychology*, 24(3), 187-201.
- Browne, T. K. (2017b). An ethics review panel for the DSM : A worthwhile challenge. *Philosophy, Psychiatry, & Psychology*, 24(3), 235-240.
- Bueter, A. (2010). Social objectivity and the problem of local epistemologies. *Analyse & Kritik*, 32(2), 213-230.

- Bueter, A. (2018). Public epistemic trustworthiness and the integration of patients in psychiatric classification. *Synthese*. doi.org/10.1007/s11229-018-01913-z
- Bueter, A. (2019). Epistemic injustice and psychiatric classification. *Philosophy of Science*, 86(5), 1064-1074.
- Bueter, A. (2019). Social epistemology and psychiatry. Dans R. Bluhm et S. Tekin (dir.), *Bloomsbury companion to philosophy of psychiatry* (p.485-503). Londres : Bloomsbury Academic.
- Callaghan, G. M., Chacon, C., Coles, C., Botts, J. et Laraway S. (2009). An empirical evaluation of the diagnostic criteria for premenstrual dysphoric disorder : Problems with sex specificity and validity. *Women & Therapy*, 32(1), 1-21.
- Callahan, G. (2009). *Between XX and XY : Intersexuality and the myth of two sexes*. Chicago : Chicago Review Press.
- Campaner, R. (2011). *Understanding mechanisms in the health sciences*. *Theoretical Medicine and Bioethics*, 32(1), 5-17.
- Caplan, P. J., McCurdy-Myers, J. et Gans, M. (1992). Should premenstrual syndrome be called a psychiatric abnormality? *Feminism & Psychology*, 2(1), 27-44.
- Caplan, P. J. (1995). *They say you're crazy : How the world's most powerful psychiatrists decide who's normal*. Caplan, MA : Addison-Wesley.
- Caplan, P. J. (2004). The debate about PMDD and Sarafem. *Women & Therapy*, 27(3-4), 55-67.
- Cartwright, S. (1851/2004). Report on the diseases and physical peculiarities of the Negro race. Dans A. L. Caplan, J. J. McCartney et D. A. Sisti (dir.), *Health, disease, and illness : Concepts in medicine* (p. 28-39). Washington : Georgetown University Press.
- Chapman, R. et Veit, W. (2020). Representing the autism spectrum. *The American Journal of Bioethics*, 20(4), 46-48.

- Clough, S. (2003). *Beyond epistemology : A pragmatist approach to feminist science studies*. Landham : Rowman and Littlefield.
- Clouser K. D., Culver C. M. et Gert B. (1981). Malady : A new treatment of disease. *The Hastings Center Report*, 11(3), 29-37.
- Collins, P. H. (1986). Learning from the outsider within : The sociological significance of Black feminist thought. *Social Problems*, 33, 14-32.
- Conrad, P. et Potter, D. (2000). From hyperactive children to ADHD adults : Observations on the expansion of medical categories. *Social Problems*, 47, 559-582.
- Conrad, P. et Slodden, C. (2013). The medicalization of mental disorder. Dans C. S. Aneshensel, J. C. Phelan et A. Bierman (dir.), *Handbook of the sociology of mental health* (p. 61-73). Dordrecht : Springer.
- Conrad, P. (2007). *The medicalization of society*. Baltimore : Johns Hopkins University Press.
- Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(2), 263-282.
- Cooper, R. (2005). *Classifying madness : A philosophical examination of the diagnostic and statistical manual of disorders*. Dordrecht : Springer.
- Cooper, R. (2007). Can it be a good thing to be deaf? *The Journal of Medicine and Philosophy*, 32(6), 563-583.
- Cooper, R. (2007). *Psychiatry and philosophy of science*. Londres : Routledge.
- Cooper, R. (2015). Must disorders cause harm? The changing stance of the DSM. Dans S. Demazeux et P. Singy (dir.), *The DSM-5 in perspective* (p. 83-96). Dordrecht : Springer.

- Cooper, R. (2017). A modest proposal. *Philosophy, Psychiatry, & Psychology*, 24(3), 207-209.
- Cooper, R. (2021). On harm. Dans L. Faucher et D. Forest (dir.), *Defining mental disorder : Jerome Wakefield and his critics* (p. 537-552). Cambridge, MA : MIT Press.
- Cosgrove, L. et Riddle, B. (2004). Gender bias and sex distribution of mental disorders in the DSM-IV-TR. Dans P. J. Caplan et L. Cosgrove (dir.), *Bias in psychiatric diagnosis* (p. 127-140). Lanham, MA : Jason Aronson.
- Cosgrove, L. et Caplan, P. J. (2004). Medicalizing menstrual distress. Dans P. J. Caplan et L. Cosgrove (dir.), *Bias in psychiatric diagnosis* (p. 221-230). Lanham, MD : Jason Aronson.
- Cosgrove, L. et Krinsky, S. (2012). A comparison of DSM-IV and DSM-5 panel members' financial associations with industry : A pernicious problem persists. *PLOS Medecine*, 9(3). doi.org/10.1371/journal.pmed.1001190
- Cosgrove, L. et Riddle, B. (2004). Gender bias and sex distribution of mental disorders in the DSM-IV-TR. Dans P. J. Caplan et L. Cosgrove (dir.), *Bias in psychiatric diagnosis* (p. 127-140). Lanham, MD : Jason Aronson.
- Cosgrove, L. et Wheeler, E. (2013a). Industry's colonization of psychiatry : Ethical and practical implications of financial conflicts of interest in the DSM-5. *Feminism & Psychology*, 23, 93-106.
- Cosgrove, L. et Wheeler, E. E. (2013b). Drug firms, the codification of diagnostic categories, and bias in 929 clinical guidelines. *The Journal of Law, Medicine & Ethics*, 41(3), 644-653.
- Cosgrove, L., Krinsky, S., Vijayaraghavan, M. et Schneider, L. (2006). Financial ties between DSM-IV panel members and the pharmaceutical industry. *Psychotherapy and Psychosomatics*, 75(3), 154-160.
- Cosgrove, L., Krinsky, S., Wheeler, E. E., Kaitz, J., Greenspan, S. B. et DiPentima, N. L. (2014). Tripartite conflicts of interest and high stakes patent extensions in the DSM-5. *Psychotherapy and Psychosomatics*, 83(2), 106-113.

- Cosmides, L. et Tooby, J. (1999). Toward an evolutionary taxonomy of treatable conditions. *Journal of Abnormal Psychology*, 108(3), 453-464.
- Cranor, C.F. (1990). Some moral issues in risk assessment. *Ethics*, 101(1), 123-143.
- Crasnow, S. L. (1993). Can science be objective? Longino's *Science as social knowledge*. *Hypatia*, 8(3), 194-201.
- Crasnow, S. (2001). Models and reality : When science tackles sex. *Hypatia*, 16(3), 138-148.
- Crasnow, S. L. (2003). Can science be objective? Dans C. Pinnick, N. Koertge et R. Almeder (dir.), *Scrutinizing feminist epistemology : An examination of gender in science* (p. 130-141). Piscataway, NJ : Rutgers University Press.
- Crasnow, S. (2006). Feminist anthropology and sociology : Issues for social science. Dans S. Turner et M. Risjord (dir.), *Handbook of the philosophy of science, volume 15 : Philosophy of anthropology and sociology* (p. 755-789). Amsterdam : Elsevier.
- Crasnow, S. (2013). Feminist philosophy of science : Values and objectivity. *Philosophy Compass*, 8(4), 413-423.
- Crasnow, S. (2018). Contemporary standpoint theory : Tensions, integrations, and extensions. Dans P. Garavaso (dir.), *The Bloomsbury companion to analytic feminism* (p. 188-211). Londres : Bloomsbury Publishing.
- Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, 68(1), 53-74.
- Craver, C. (2005). Functions and mechanisms in contemporary neuroscience. Dans P. Poirier, L. Faucher, E. Racine et E. Ennan (dir.), *Des neurones à la conscience : Neurophilosophie et philosophie des neurosciences*. Bruxelles : De Boeck Université.

- Craver, C. F. (2007). *Explaining the brain : Mechanisms and the mosaic unity of neuroscience*. New York : Oxford University Press
- Craver, C. F. (2013). Functions and mechanisms : A perspectivalist view. Dans P. Huneman (dir.), *Functions : Selection and mechanisms* (p. 133-158). Dordrecht : Springer.
- Craver, C. F. et L. Darden (2013). *In search of mechanisms : Discoveries across the life sciences*. Chicago : University of Chicago Press.
- Craver, C. et Tabery, J. (2019). Mechanisms in Science. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>
- Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA : Bradford/ MIT Press.
- Cuthbert, B. N. (2015). Research domain criteria : Toward future psychiatric nosologies. *Dialogues in Clinical Neuroscience*, 17(1), 89-97.
- Daston, D. et Galison, P. (2007). *Objectivity*. Cambridge, MA : MIT Press.
- Davies, P. S. (2001). *Norms of nature : Naturalism and the nature of functions*. Cambridge, MA : MIT Press.
- De Block, A. et Sholl, J. (2021). Harmless dysfunctions and the problem of normal variation. Dans L. Faucher et D. Forest (dir.), *Defining mental disorders : Jerome Wakefield and his critics* (p. 495-510). Cambridge, MA : MIT Press.
- Decker, H. S. (2007). How Kraepelinian was Kraepelin? How Kraepelinian are the neo-Kraepelinians? From Emil Kraepelin to *DSM-III*. *History of Psychiatry*, 18, 337-360.
- Decker, H. S. (2013). *The making of DSM-III : A diagnostic manual's conquest of American psychiatry*. Oxford : Oxford University Press.

- Demazeux, S. (2015). The function debate and the concept of mental disorder. Dans • P. Huneman, G. Lambert et M. Silberstein (dir.), *Classification, disease and evidence* (p. 63-91). Springer : Dordrecht.
- Demazeux, S. (2021). *DSM in the light of HDA (and conversely)*. Dans L. Faucher et D. Forest (dir.), *Defining mental disorders : Jerome Wakefield and his critics* (p. 3-26). Cambridge, MA : MIT Press.
- DeVito, S. (2000). On the value-neutrality of the concepts of health and disease : Unto the breach again. *The Journal of Medicine and Philosophy*, 25(5), 539-567.
- De Vreese, L. (2017). How to proceed in the disease concept debate? A pragmatic approach. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 42(4), 424-446.
- Di Giulio, G. G. et Reissing, E. D. (2006). Premenstrual Dysphoric Disorder : Prevalence, Diagnostic Considerations, and Controversies. *Journal of Psychosomatic Obstetrics & Gynecology*, 27(4), 201-210.
- Dorato, M. (2004). Epistemic and non-epistemic values in science. Dans P. Machamer et G. Wolters (dir.), *Science, values, and objectivity* (p. 52-77). Pittsburgh : University of Pittsburgh Press.
- Double, D. B. (2006). Historical perspective on anti-psychiatry. Dans D. B. Double (dir.), *Critical psychiatry : The limit of madness* (p. 19-39). Londres : Palgrave Macmillan.
- Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, 67(4), 559-579.
- Douglas, H. (2003). The moral responsibilities of scientists (tensions between autonomy and responsibility). *American Philosophical Quarterly*, 40(1), 59-68.
- Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, 138(3), 453-473.

- Douglas, H. (2007). Rejecting the ideal of value-free science. Dans H. Kincaid, J. Dupré et A. Wylie (dir.), *Value-free science? Ideals and illusions* (p. 120-141). New York : Oxford University Press.
- Douglas, H. (2008). The role of values in expert reasoning. *Public Affairs Quarterly*, 22(1), 1-18.
- Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh : University of Pittsburgh Press.
- Doust, J., Walker, M. et Rogers, W. A. (2017). Current dilemmas in defining the boundaries of disease. *Journal of Medicine and Philosophy*, 42(4), 350-366.
- Dowrick, C. et Frances, A. (2013). Medicalising unhappiness : New classification of depression risks more patients being put on drug treatment from which they will not benefit. *British Medical Journal*.
doi.org/10.1371/journal.pmed.1001190
- Dupré, J. (2007). Fact and value. Dans H. Kincaid, J. Dupré et A. Wylie (dir.), *Value-free science? Ideals and illusions* (p. 27-41). New York : Oxford University Press.
- Eckardt, B. V. et Poland, J. S. (2004). Mechanism and explanation in cognitive neuroscience. *Philosophy of Science*, 71(5), 972-984.
- Elliott, K. C. (2011). *Is a little pollution good for you? Incorporating societal values in environmental research*. New York : Oxford University Press.
- Engelhardt, H. T. (1974). The disease of masturbation : Values and the concept of disease. *Bulletin of the History of Medicine*, 48(2), 234-248.
- Engelhardt, H. T., Jr. (1975). The concepts of health and disease. Dans H. T. Engelhardt, Jr. et S. F. Spicker (dir.), *Evaluation and explanation in the biological sciences* (p. 125-141). Dordrecht : D. Reidel.

- Erciyès, B. (2015). *Situating objectivity : A feminist conceptualization* (thèse de doctorat). York University, Toronto.
- Ereshefsky, M. (2009). Defining “health” and “disease”. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 40(3), 221-227.
- Faucher, L. et Blanchette, I. (2011). Fearing new dangers : Phobias and the cognitive complexity of human emotions. Dans A. De Block et P. Adriaens (dir.), *Maladapting minds : Philosophy, psychiatry, and evolutionary theory* (p. 33-64). New York : Oxford University Press.
- Faucher, L. et Forest, D. (2021). *Defining mental disorders : Jerome Wakefield and his critics*. Cambridge, MA : MIT Press.
- Fausto-Sterling, A. (2000). *Sexing the body : Gender politics and the construction of sexuality*. New York : Basic Books.
- Feyerabend, P. (1962). Explanation, reduction and empiricism. Dans H. Feigl et G. Maxwell (dir.), *Scientific explanation, space, and time* (p. 28-97). Minneapolis : University of Minneapolis Press.
- First, M. B. et Wakefield, J. C. (2013). Diagnostic criteria as dysfunction indicators : Bridging the chasm between the definition of mental disorder and diagnostic criteria for specific disorders. *The Canadian Journal of Psychiatry*, 58(12), 663-669.
- Forster, E. M. (1910/1997). *Howards end*. New York : Macmillan.
- Frances, A. J. et Widiger, T. (2012). Psychiatric diagnosis : Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, 8, 109-130.
- Freedman, K. L. (2009). Diversity and the fate of objectivity. *Social Epistemology*, 23(1), 45-56.

- Fulford, K. W. M. (1989). *Moral theory and medical practice*. Cambridge : Cambridge University Press.
- Fulford, K. W. M. (1991). Evaluative delusions : Their significance for philosophy and psychiatry. *British Journal of Psychiatry*, 159, 108-112.
- Fulford, K. W. M. (1999). Nine variations and a coda on the theme of an evolutionary definition of dysfunction. *Journal of Abnormal Psychology*, 108(3), 412-420.
- Fulford, K. W. M. (2000). Teleology without tears : Naturalism, neo-naturalism, and evaluationism in the analysis of function statements in biology (and a bet on the twenty-first century). *Philosophy, Psychiatry, & Psychology*, 7(1), 77-94.
- Fulford, K. W. M. (2001). "What is (mental) disease?" : An open letter to Christopher Boorse. *Journal of Medical Ethics*, 27(2), 80-85.
- Fulford, K. W. M. (2005). Values in psychiatric diagnosis : Developments in policy, training and research. *Psychopathology*, 38, 171-176.
- Fuller, S. (1988). *Social epistemology*. Bloomington, IN : Indiana University Press.
- Gamwell, L. et Tomes. N. (1995). *Madness in America : Cultural and medical perspectives of mental illness before 1914*. Ithaca, NY : Cornell University Press.
- Garson, J. R. (2006). *Psychiatric disorders and biological dysfunctions : Some philosophical questions concerning psychiatry*. (Thèse de doctorat non publiée). University of Texas at Austin.
- Garson, J. (2013). The functional sense of mechanism. *Philosophy of Science*, 80(3), 317-333.
- Garson, J. (2016). *A critical overview of biological functions*. Cham : Springer International Publishing.
- Garson, J. (2019). There are no ahistorical theories of function. *Philosophy of Science*, 86(5), 1146-1156.

- Glackin, S. (2019). Grounded disease : Constructing the social from the biological in medicine. *The Philosophical Quarterly*, 69, 258-276.
- Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 49-71.
- Goffman, E. (1961). *Asylums : Essays on the social situations of mental patients and other inmates*. Oxford : Doubleday.
- Gold, I. et Kirmayer, L. J. (2007). Cultural psychiatry on Wakefield's procrustean bed. *World Psychiatry*, 6(3), 165-166.
- Goldman, A. et Blanchard, T. (2018). Social Epistemology. Dans E. N. Zalta (dir.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/fall2019/entries/epistemology-social/>
- Grob, G. N. (1991). Origins of *DSM-I* : A Study in Appearance and Reality. *American Journal of Psychiatry*, 148(4), 421-431.
- Grob, G. N. (1994b). *The mad among us : A history of the care of America's mentally ill*. Cambridge, MA : Harvard University Press.
- Grob, G. N. (2011). The attack of psychiatric legitimacy in the 1960s : Rhetoric and reality. *Journal of the History of the Behavioral Sciences*, 47(4), 398-416.
- Gross, P. R. et Levitt, N. (1997). *Higher superstition : The academic left and its quarrels with science*. Baltimore : Johns Hopkins University Press.
- Gureje, O. et Stein, D. J. (2012). Classification of mental disorders : The importance of inclusive decision-making. *International Review of Psychiatry*, 24(6), 606-612.
- Haack, S. (1993). Epistemological reflections of an old feminist. *Reason Papers*, 18, 31-43.

- Haack, S. (2003). Knowledge and propaganda. Dans C. Pinnick, N. Koertge et R. Almeder (dir.), *Scrutinizing feminist epistemology : An examination of gender in science* (p. 7-19). Piscataway, NJ : Rutgers University Press.
- Habermas, J. (1992/1996). *Between facts and norms : Contributions to a discourse theory of law and democracy*. (W. Rehg, trad.). Cambridge, MA : MIT Press.
- Haely, K.C. (2008). *Objectivity in the feminist philosophy of science*. Londres : A&C Black.
- Hale, N. G. (1995). *The rise and crisis of psychoanalysis in America : Freud and the Americans, 1917-1985*. New York : Oxford University Press.
- Hankinson-Nelson, L. (1990). *Who knows? From Quine to a feminist empiricism*. Philadelphia : Temple University Press.
- Haque, O. S., Lu, A., Wu, D., Cosgrove, L. et Bursztajn, H. J. (2014). Curing financial conflicts of interest in psychiatric professional organizations. Dans J. Z. Sadler, B. Fulford et C. W. van Staden (dir.), *The Oxford handbook of psychiatric ethics*. Oxford : Oxford University Press.
- Haraway, D. (1988). Situated knowledges : The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575-599.
- Hardcastle, V. G. (1999). Understanding functions. Dans V. G. Hardcastle (dir.), *Where biology meets psychology* (p. 27-43). Cambridge, MA : MIT Press.
- Hardcastle, V. G. (2002). On the normativity of functions. Dans A. R. Ariew, R. Cummins et M. Perlman (dir.), *Functions : New essays in the philosophy of psychology and biology* (p. 144-156). Oxford : Oxford University Press.
- Harding, S. G. (1986). *The science question in feminism*. Ithaca, NY : Cornell University Press.
- Harding, S. (2015). *Objectivity and diversity : Another logic of scientific research*. Chicago : University of Chicago Press.

- Heinrichs, R. W. (2001). *In search of madness : Schizophrenia and neuroscience*. New York : Oxford University Press.
- Hekman, S. (1997). Truth and method : Feminist standpoint theory revisited. *Signs : Journal of Women in Culture and Society*, 22(2), 341-365.
- Hicks, D. (2011). Is Longino's conception of objectivity feminist? *Hypatia*, 26(2), 333-351.
- Hjørland, B. (2016). The paradox of atheoretical classification. *Knowledge Organization*, 43(5), 313-323.
- Horwitz, A. V. et Wakefield, J. C. (2007). *The loss of sadness : How psychiatry transformed normal sorrow into depressive disorder*. New York : Oxford University Press.
- Houts, A. C. (2001). Harmful dysfunction and the search for value neutrality in the definition of mental disorder : Response to Wakefield, part 2. *Behaviour Research and Therapy*, 39(9), 1099-1132.
- Howalter, E. (1987). *The female malady*. Londres : Virago.
- Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, 8(9), 725-732.
- Hyman, S. E. (2010). The diagnosis of mental disorders : The problem of reification. *Annual Review of Clinical Psychology*, 6, 155-179.
- Intemann, K. (2005). Feminism, underdetermination, and values in science. *Philosophy of Science*, 72(5), 1001-1012.
- Intemann, K. (2008). Increasing the number of feminist scientists : Why feminist aims are not served by the underdetermination thesis. *Science & Education*, 17(10), 1065-1079.
- Intemann, K. (2010). 25 years of feminist empiricism and standpoint theory : Where are we now? *Hypatia*, 25(4), 778-796.

- Intemann, K. (2011). Diversity and dissent in science : Does democracy always serve feminist aims? Dans H. E. Grasswick (dir.), *Feminist epistemology and philosophy of science* (p. 111-132). Dordrecht : Springer.
- Intemann, K. (2016). Feminist standpoint. Dans L. Disch et M. Hawkesworth (dir.), *The Oxford handbook of feminist theory* (p. 261-282). Oxford : Oxford University Press.
- Intemann, K. (2017). Feminism, values, and the bias paradox. Dans K. C. Elliott et D. Steel (dir.), *Current controversies in values and science* (p. 130-131). New York : Routledge.
- Jablensky, A. (2007). Living in a Kraepelinian world : Kraepelin's impact on modern psychiatry. *History of Psychiatry, 18*, 381-388.
- Jablensky, A. (2007). Does psychiatry need an overarching concept of "mental disorder"? *World Psychiatry, 6*(3), 157-158.
- Janack, M. (2002). Dilemmas of objectivity. *Social Epistemology, 16*(3), 267-281.
- Karin, B. (dir.) (2010). *Environmental politics and deliberative democracy : Examining the promise of new modes of governance*. Northampton, MA : Edward Elgar Publishing.
- Kellert, S., Longino, H. et Waters, K. (2006). Introduction : The pluralist stance. Dans S. Kellert, H. Longino et K. Waters (dir.), *Scientific pluralism* (p. vii-xxix). Minneapolis : University of Minnesota Press.
- Kendell, R. E. (1986). What are mental disorders? Dans M. Freedman, R. Brotman et I. Silverman (dir.), *Issues in psychiatric classification* (p. 23-45). New York : Human Sciences Press.
- Kendler, K. S., Appelbaum, P. S., Bell, C. C., Fulford, K. W. M., Ghaemi, S. N., Schaffner, K. F., ... Sadler, J. Z. (2008). Issues for DSM-V : DSM-V should include a conceptual issues work group. *American Journal of Psychiatry, 165*(2), 174-175.

- Kendler, K. S. et Solomon, M. (2016). Expert consensus v. evidence-based approaches in the revision of the DSM. *Psychological Medicine*, 46(11), 2255-2262.
- Kincaid, H. (2007). Contextualist morals and science. Dans H. Kincaid, J. Dupré et A. Wylie (dir.), *Value-free science : Ideals and illusions?* (p. 218-238). Oxford : Oxford University Press.
- Kinghorn, W. (2013). The biopolitics of defining “mental disorder”. Dans J. Parris et J. Phillips (dir.), *Making the DSM-5* (p. 47-61). New York : Springer.
- Kingma, E. (2007). What is it to be healthy? *Analysis*, 67(2), 128-133.
- Kingma, E. (2010). Paracetamol, poison, and polio : Why Boorse’s account of function fails to distinguish health and disease. *The British Journal for the Philosophy of Science*, 61(2), 241-264.
- Kingma, E. (2014). Naturalism about health and disease : Adding nuance for progress. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 39(6), 590-608.
- Kingma, E. (2016). Situation-specific disease and dispositional function. *The British Journal for the Philosophy of Science*, 67(2), 391-404.
- Kirk, S. A. et Kutchins, H. (1992). *The selling of DSM : The rhetoric of science in psychiatry*. New York : Routledge.
- Kirmayer, L. J. et Young, A. (1999). Culture and context in the evolutionary concept of mental disorder. *Journal of Abnormal Psychology*, 108(3), 446-452.
- Kitcher, P. (1993). Function and design. *Midwest Studies in Philosophy*, 18, 379-397.
- Koskinen, I. (2017). Where is the epistemic community? On democratisation of science and social accounts of objectivity. *Synthese*, 194(12), 4671-4686.
- Koskinen, I. (2018). Defending a risk account of scientific objectivity. *The British Journal for the Philosophy of Science*.

- Kuhn, T. S. (1970). *The structure of scientific revolutions*. (2^e édition). Chicago : University of Chicago Press.
- Kupfer, David J., First, Michael B. et Regier, Darrel A. (dir.). (2002). *A research agenda for DSM-5*. Washington : American Psychiatric Association.
- Kutchins, H. et Kirk, S. A. (1997). *DSM : The psychiatric bible and the creation of mental disorders*. New York : The Free Press.
- Lane, C. (2007). *Shyness : How normal behavior became a sickness*. New Haven, CT : Yale University Press.
- Lange, M. (2007). The end of disease. *Philosophical Topics*, 35, 265-292.
- Lemoine, M. (2013). Defining disease beyond conceptual analysis : An analysis of conceptual analysis in philosophy of medicine ». *Theoretical Medicine and Bioethics*, 34(4), 309-325.
- Lemoine, M. (2014). Trois conceptions sémantiques des théories en médecine. *Lato Sensu, Revue de la Société de philosophie des sciences*, 1(1), 1-11.
- Lemoine, M. (2017). *Introduction à la philosophie des sciences médicales*. Paris : Hermann.
- Lemoine, M. (2021) Is the dysfunction component of the “harmful dysfunction analysis” stipulative? Dans L. Faucher et D. Forest, *Defining mental disorder : Jerome Wakefield and his critics* (p. 199-212). Cambridge, MA : MIT Press.
- Lemoine, M. et Giroux, É. (2016). Is Boorse’s biostatistical theory of health naturalistic? Dans Élodie Giroux (dir.), *Naturalism in the philosophy of health* (p. 19-38). s.l. (Suisse) : Springer International Publishing.
- Lexchin, J. et O’Donovan, O. (2010). Prohibiting or “managing” conflict of interest? A review of policies and procedures in three European drug regulation agencies. *Social Science and Medicine*, 70, 643-647.

- Lilienfeld, S. O. et Marino, L. (1995). Mental disorder as a Roschian concept : A critique of Wakefield's « harmful dysfunction » analysis. *Journal of Abnormal Psychology*, 104(3), 411-420.
- Lilienfeld, S. O. et Marino, L. (1999). Essentialism revisited : Evolutionary theory and the concept of mental disorder. *Journal of Abnormal Psychology*, 108(3), 400-411.
- Lloyd, E. (1995). Objectivity and the double standard for feminist epistemologies. *Synthese*, 104(3), 351-381.
- Longino, H. (1983). Beyond “bad science” : Skeptical reflections on the value-freedom of scientific inquiry. *Science, Technology & Human Values*, 8(1), 7-17.
- Longino, H. E. (1990). *Science as social knowledge : Values and objectivity in scientific inquiry*. Princeton, NJ : Princeton University Press.
- Longino, H. (1993). Subjects, power, and knowledge : Description and prescription in feminist philosophies of science. Dans L. Alcof et E. Potter (dir.), *Feminist epistemologies* (p. 101-120). New York : Routledge.
- Longino, H. E. (1995). Gender, politics, and the theoretical virtues. *Synthese*, 104(3), 383-397.
- Longino, H. E. (2002). *The fate of knowledge*. Princeton, NJ : Princeton University Press.
- Loring, M. et Powell, B. (1988). Gender, race, and DSM-III : A study of the objectivity of psychiatric diagnostic behavior. *Journal of Health and Social Behavior*, 29(1), 1-22.
- Lunbeck, E. (1994). *The psychiatric persuasion : Knowledge, gender, and power in modern America*. Princeton, NJ : Princeton University Press.

- MacDougall, D. R. (2020). The ends of medicine and the experience of patients. *The Journal of Medicine and Philosophy : A Forum for Bioethics and Philosophy of Medicine*, 45(2), 129-144.
- Machamer, P., Darden, L. et Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1-25.
- MacKinnon, C. A. (1989). *Toward a feminist theory of the state*. Cambridge, MA : Harvard University Press.
- Martí, J. L. (2006). The epistemic conception of deliberative democracy defended. Dans S. Besson et J. L. Martí (dir.), *Deliberative democracy and its discontents* (p. 27-56). Aldershot : Ashgate.
- Martin, E. (1991). The egg and the sperm : How science has constructed a romance based on stereotypical male-female roles. *Signs : Journal of Women in Culture and Society*, 16(3), 485-501.
- Martin, N. (2001). Feminist bioethics and psychiatry. *The Journal of Medicine and Philosophy*, 26(4), 431-441.
- Matheson, A. (2008). Corporate science and the husbandry of scientific and medical knowledge by the pharmaceutical industry. *BioSocieties*, 3(4), 355-382.
- Mayes, R. et Horwitz, A. V. (2005). DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences*, 41, 249-267.
- Mcallister, J. W. (2017). Relativism. Dans W. H. Newton-Smith (dir.), *A Companion to the philosophy of science* (p. 405-407). Oxford : Blackwell Publishers.
- McCaffrey, J. B. (2015). The brain's heterogeneous functional landscape. *Philosophy of Science*, 82(5), 1010-1022.
- McCoy, M. S., Liu, E. Y., Lutz, A. S. et Sisti, D. (2020). Ethical advocacy across the autism spectrum : Beyond partial representation. *The American Journal of Bioethics*, 20(4), 13-24.

- McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A*, 16(3), 247-273.
- McNally, R. J. (2001). On Wakefield's harmful dysfunction analysis of mental disorder. *Behaviour Research and Therapy*, 39(3), 309-314.
- Megone, C. (1998). Aristotle's function argument and the concept of mental illness. *Philosophy, Psychiatry, & Psychology*, 5(3), 187-201.
- Megone, C. (2000). Mental illness, human function, and values. *Philosophy, Psychiatry, & Psychology*, 7(1), 45-65.
- Metzl, J. M. (2009). *The protest psychosis : How schizophrenia became a black disease*. Boston : Beacon Press.
- Millikan, R. G. (1989a). An ambiguity in the notion "function". *Biology and Philosophy*, 4(2), 172-176.
- Millikan, R. G. (1989b). In defense of proper functions. *Philosophy of Science*, 56, 288-302.
- Moon, M. R. et Khin-Maung-Gyi, F. (2009). The history and role of institutional review boards. *Virtual Mentor*, 11(4), 311-316.
- Moynihan, R. N., Cooke, G. P., Doust, J. A., Bero, L., Hill, S. et Glasziou, P. P. (2013). Expanding disease definitions in guidelines and expert panel ties to industry : A cross-sectional study of common conditions in the United States. *PLOS Medicine*, 10(8).
- Murphy, D. et Woolfolk, R. L. (2000a). Conceptual analysis versus scientific understanding : An assessment of Wakefield's folk psychiatry. *Philosophy, Psychiatry, & Psychology*, 7(4), 271-293.
- Murphy, D. (2005). Can evolution explain insanity? *Biology and Philosophy*, 20(4), 745-766.
- Murphy, D. (2006). *Psychiatry in the scientific image*. Cambridge, MA : MIT Press.

- Murphy, D. (2010). Complex mental disorders : Representation, stability and explanation. *European Journal of Analytic Philosophy*, 6(1), 28-42.
- Murphy, D. (2011). Conceptual foundations of biological psychiatry. Dans F. Gifford (dir.), *Philosophy of Medicine* (p. 425-451). North Holland : Elsevier.
- Murphy, D. (2020). Concepts of disease and health. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/spr2020/entries/health-disease/>
- Murphy, D. (2021). Function and dysfunction. Dans L. Faucher et D. Forest (dir.), *Defining mental disorders : Jerome Wakefield and his critics* (p. 251-266). Cambridge, MA : MIT Press.
- Murphy, D. et Woolfolk, R. L. (2000). The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, & Psychology*, 7(4), 241-252.
- Murphy, D. et Stich, S. (2000). Darwin in the madhouse : Evolutionary psychology and the classification of mental disorders. Dans P. Carruthers et A. Chamberlain (dir.), *Evolution and the human mind : Modularity, language and meta-cognition* (p. 62-92). Cambridge : Cambridge University Press.
- Musschenga, A. W., Van der Steen, Wim J. et Ho, Vincent K. Y. (2010). The Business of Drug Research : A Mixed Blessing. Dans H. Radder (dir.), *The commodification of academic science* (p. 110-131). Pittsburgh : Pittsburgh University Press.
- Nagel, E. et Hawkins, D. (1961). The structure of science. *American Journal of Physics*, 29(10), 716-716.
- Nagel, T. (1986). *The view from nowhere*. New York : Oxford University Press.
- Narrow, W. E. et Kuhl, E. A. (2010). Clinical significance and disorder thresholds in DSM-5. Dans D. A. Regier, W. E. Narrow, E. A. Kuhl et D. J. Kupfe (dir.), *The conceptual evolution of DSM-5* (p. 147-162). Arlington, VA: American Psychiatric Publishing.

- Neander, K. (1991). Functions as selected effects : The conceptual analyst's defense. *Philosophy of Science*, 58(2), 168-184.
- Niiniluoto, I. (1996). The relativism question in feminist epistemology. Dans L. H. Nelson et J. Nelson (dir.), *Feminism, science, and the philosophy of science* (p. 139-157). Dordrecht : Springer.
- Nissim-Sabat, M. (2009). *Neither victim nor survivor : Thinking toward a new humanity*. Plymouth : Lexington Books.
- Nissim-Sabat, M. (2013). Race and gender in philosophy of psychiatry : Science, relativism, and phenomenology. Dans K. W. M. Fulford, M. Davies, G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini et T. Thornton (dir.), *The Oxford Handbook of Philosophy and Psychiatry* (p. 139-158). Oxford : Oxford University Press.
- Nordby, H. (2006). The analytic-synthetic distinction and conceptual analyses of basic health concepts. *Medicine, Health Care and Philosophy*, 9(2), 169-180.
- Nordenfelt, L. (1995). *On the nature of health : An action-theoretic approach*. Dordrecht : Springer.
- Nordenfelt, L. (2001). *Health, science, and ordinary language*. Amsterdam : Rodopi.
- Offman, A. et Kleinplatz, P. J. (2004). Does PMDD belong in the DSM? Challenging the medicalization of women's bodies. *Canadian Journal of Human Sexuality*, 13(1), 17-27.
- Organisation mondiale de la Santé (1946). Préambule à la constitution de l'Organisation mondiale de la Santé, tel qu'adopté par la Conférence internationale sur la Santé, New York, 19 juin-22 juillet 1946; signé le 22 juillet 1946 par les représentants de 61 États. Dans *Actes officiels de l'Organisation mondiale de la Santé*, 2, 100. Genève : OMS.
- Phillips, J., Frances, A., Cerullo, M. A., Chardavoyne, J., Decker, H. S., First, M. B., ... Zachar, P. (2012). The six most essential questions in psychiatric

diagnosis : A pluralogue. Part 1 : Conceptual and definitional issues in psychiatric diagnosis. *Philosophy, Ethics and Humanities in Medicine*, 7(3), 1-29.

Piccinini, G. et Craver, C. F. (2011). Integrating psychology and neuroscience : Functional analyses as mechanism sketches. *Synthese*, 183(3), 283-311.

Pilecki, B. C., Clegg, J. W. et McKay, D. (2011). The influence of corporate and political interests on models of illness in the evolution of the DSM. *European Psychiatry*, 26(3), 194-200.

Pinnick, C. L., Koertge, N. et Almeder, R. F. (dir.). (2003). *Scrutinizing feminist epistemology : An examination of gender in science*. Piscataway, NJ : Rutgers University Press.

PLOS Medicine Editors (2012). Does conflict of interest disclosure worsen bias? *PLOS Medicine*, 9(4). doi.org/10.1371/journal.pmed.1001210

Plutynski, A. (2018). *Explaining cancer : Finding order in disorder*. Oxford : Oxford University Press.

Poland, J., Von Eckardt, B. et Spaulding, W. (1994). Problems with the DSM approach to classifying psychopathology. Dans G. Graham et L. G. Stephens (dir.), *Philosophical psychopathology* (p. 235-260). Cambridge (Mass.) : MIT Press.

Porter, R. (2002). *Madness : A brief history*. Oxford : Oxford University Press.

Porter, D. (2017). Ontological assumptions, a biopsychosocial approach, and patient participation : Moving toward an ethically legitimate science of psychiatric nosology. *Philosophy, Psychiatry, & Psychology*, 24(3), 223-226.

Potter, E. (2006). *Feminism and philosophy of science : An introduction*. New York : Routledge.

Potter, N. N. (2006). What is manipulative behavior, anyway? *Journal of Personality Disorders*, 20, 139-156.

- Potter, N. N. (2008). The problem with too much anger : A philosophical approach to understanding anger in borderline personality disordered patients. Dans C. Charland et P. Zachar (dir.), *Fact and value in emotion* (p. 53-86). Amsterdam : John Benjamins.
- Potter, N. N. (2012). Mad, bad, or virtuous? The moral, cultural, and pathologizing features of defiance. *Theory & Psychology*, 22(1), 23-45.
- Potter, N. N. (2014). Oppositional defiant disorder : Cultural factors that influence interpretations of defiant behavior and their social and scientific consequences. Dans H. Kincaid et J. A. Sullivan (dir.), *Classifying psychopathology : Mental kinds and natural kinds* (p. 175). Cambridge, MA : MIT Press.
- Pouncey, C. et Merz, J. F. (2017). Browne's external DSM ethical review panel : That dog won't hunt. *Philosophy, Psychiatry, & Psychology*, 24(3), 227-230.
- Powell, R. et Scarffe, E. (2019a). Rethinking "disease" : A fresh diagnosis and a new philosophical treatment. *Journal of Medical Ethics*, 45(9), 579-588.
- Powell, R. et Scarffe, E. (2019b). Rehabilitating "disease" : Function, value, and objectivity in medicine. *Philosophy of Science*, 86(5), 1168-1178.
- Proctor, J. D. (1998). The social construction of nature : Relativist accusations, pragmatist and critical realist responses. *Annals of the Association of American Geographers*, 88(3), 352-376.
- Proctor, R. (2000). Nazi science and Nazi medical ethics : Some myths and misconceptions. *Perspectives in Biology and Medicine*, 43(3), 335-346.
- Putnam, H. (2002). *The collapse of the fact/value dichotomy*. Cambridge, MA : Harvard University Press.
- Radden, J. (2019). Mental Disorder (Illness). Dans E. N. Zalta (dir.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/win2019/entries/mental-disorder/>

- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA : Harvard University Press.
- Regier, D. A., Kuhl, E. A., Kupfer, D. J. et McNulty, J. P. (2010). Patient involvement in the development of DSM-V. *Psychiatry : Interpersonal and Biological Processes*, 73(4), 308-310.
- Regier, D. A., Narrow, W. E., First, M. B. et Marshall, T. (2002). The APA classification of mental disorders : Future perspectives. *Psychopathology*, 35(2-3), 166-170.
- Regier, D. A., Narrow, W. E., Kuhl, E. A., Kupfer, D. J. et American Psychopathological Association (2011). *The conceptual evolution of DSM-5*. Arlington, VA : American Psychiatric Publishing.
- Reiss, J. et Sprenger, J. (2017). Scientific Objectivity. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/>
- Richards, J. R. (1995). Why feminist epistemology isn't (and the implications for feminist jurisprudence). *Legal Theory*, 1(4), 365-400.
- Roberts, D. (2013). Thick concepts. *Philosophy Compass*, 8(8), 677-688.
- Richters, J. E. et Hinshaw, S. P. (1999). The abduction of disorder in psychiatry. *Journal of Abnormal Psychology*, 108(3), 438-445.
- Roe, K. et Murphy, D. (2011). Function, dysfunction, adaptation? Dans P. Adriaens et A. Block (dir.), *Maladapting minds* (p. 216-237). New York : Oxford University Press.
- Rogers, W. A. et Walker, M. J. (2017). The line-drawing problem in disease definition. *Journal of Medicine and Philosophy*, 42(4), 405-423.
- Rolin, K. (2009a). Standpoint theory as a methodology for the study of power relations. *Hypatia*, 24(4), 218-226.

- Rolin, K. (2009b). Scientific knowledge : A stakeholder theory. Dans J. Van Bouwel (dir.), *The social sciences and democracy* (p. 62-80). Basingstoke : Palgrave Macmillan.
- Rolin, K. (2011). Contextualism in feminist epistemology and philosophy of science. Dans H. E. Grasswick (dir.), *Feminist epistemology and philosophy of science* (p. 25-44). Dordrecht : Springer.
- Rolin, K. (2017). Can social diversity be best incorporated into science by adopting the social value management ideal? Dans K. C Elliott et D. Steel (dir.), *Current controversies in values and science* (p. 113-129). New York : Routledge.
- Ross, P. A. (2002). The fact value dichotomy in demarcating disorder. *Philosophy, Psychiatry, & Psychology*, 14(2), 107-109.
- Rounsaville, B. J., Alarcón, R. D., Andrews, G., Jackson, J. S., Kendell, R. E. et Kendler, K. (2002). Basic nomenclature issues for DSM-V. Dans D. J. Kupfer, M. B. First et D. S. Regier (dir.), *A research agenda for DSM-V* (p. 1-30). Washington : American Psychiatric Association.
- Ruphy, S. (2015). Rôle des valeurs en science : Contributions de la philosophie féministe des sciences. *Écologie & politique*, 2, 46-47.
- Ruphy, S. (2006). “Empiricism all the way down” : A defence of the value-neutrality of science in response to Helen Longino’s contextual empiricism. *Perspectives on Science*, 14(2), 189-214.
- Rysiew, P. (2016). Epistemic Contextualism. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/win2016/entries/contextualism-epistemology/>
- Sadegh-Zadeh, K. (2000). Fuzzy health, illness, and disease. *Journal of Medicine and Philosophy*, 25(5), 605-638.

- Sadegh-Zadeh, K. (2008). The prototype resemblance theory of disease. *Journal of Medicine and Philosophy*, 33(2), 106-139.
- Sadler, J. Z. et Agich, G. J. (1995). Diseases, functions, values, and psychiatric classification. *Philosophy, Psychiatry, & Psychology*, 2(3), 219-231.
- Sadler, J. Z. (2002). Values in developing psychiatric classifications : A proposal for DSM-V. Dans J. Z. Sadler (dir.), *Descriptions and prescriptions : Values, mental disorders, and the DSMs* (p. 301-322). Baltimore : John Hopkins University Press.
- Sadler, J. Z. et Fulford, B. (2004). Should patients and their families contribute to the DSM-V process? *Psychiatric Services*, 55(2), 133-138.
- Sandberg, D. E., Pasterski, V. et Callens, N. (2017). Introduction to the special section : Disorders of sex development. *Journal of Pediatric Psychology*, 42(5), 487-495.
- Sarbin, T. R. (1967). On the futility of the proposition that some people be labeled "mentally ill". *Journal of Consulting Psychology*, 31(5), 447-453.
- Schaffner, K. F. (1986). Exemplar reasoning about biological models and diseases : A relation between the philosophy of medicine and philosophy of science. *The Journal of medicine and philosophy*, 11(1), 63-80.
- Schaffner, K. F. (1993). *Discovery and explanation in biology and medicine*. Chicago : University of Chicago Press.
- Scheibe, K. E. et Barrett, F. J. (2016). Sarbin's way : Overcoming mentalism and mechanism in psychology. *Theory & Psychology*, 26(4), 516-539.
- Schmitt, F. F. (dir.). (1994). *Socializing epistemology : The social dimensions of knowledge*. Lanham, MA : Rowman & Littlefield.
- Schramme, T. (2010). Can we define mental disorder by using the criterion of mental dysfunction? *Theoretical Medicine and Bioethics*, 31(1), 35-47.

- Schuur, R. (2019). Mental health and illness : Past debates and future directions. Dans S. Tekin et R. Bluhm (dir.), *The Bloomsbury Companion to Philosophy of Psychiatry* (p. 527-541). Londres : Bloomsbury Academic.
- Schwartz, P. H. (2004). An alternative to conceptual analysis in the function debate : Introduction ». *The Monist*, 87(1), 136-153.
- Schwartz, Peter H. (2007). Defining dysfunction : Natural selection, design, and drawing a line. *Philosophy of Science*, 74(3), 364-385.
- Schwartz, P. H. (2007). Decision and discovery in defining “disease”. Dans H. Kincaid et J. McKittrick (dir.), *Establishing medical reality : Essays in the metaphysics and epistemology of biomedical science* (p. 47-63). Heidelberg : Springer Netherlands.
- Sedgwick, P. (1982). Psycho politics : Laing, Foucault, Goffman, Szasz, and the future of mass psychiatry. Cambridge : Harper & Row.
- Shaw, C. et Proctor, G. (2005). Women at the margins : A critique of the diagnosis of borderline personality disorder. *Feminism & Psychology*, 15(4), 483-490.
- Shields, S. (1975). Functionalism, darwinism, and the psychology of women. *American Psychologist*, 30(7), 739-754.
- Shields, S. A. (1982). The variability hypothesis : The history of a biological model of sex differences in intelligence. *Signs : Journal of Women in Culture and Society*, 7(4), 769-797.
- Shields, S. A. (2007). Passionate men, emotional women : Psychology constructs gender difference in the late 19th century. *History of Psychology*, 10(2), 92.
- Shields, K. et Beversdorf, D. (2020). A dilemma for neurodiversity. *Neuroethics*. <https://doi.org/10.1007/s12152-020-09431-x>
- Shorter, E. (1997). A history of psychiatry. New York : John Wiley and Sons.

- Shorter, E. (2013). The history of the DSM. Dans J. Paris et J. Phillips (dir.), *Making the DSM-5 : Concepts and controversies* (p. 3-19). New York : Springer.
- Silberman, S. (2015). *NeuroTribes : The legacy of autism and the future of neurodiversity*. New York : Avery.
- Sisti, D. et Johnson, R. (2015). Revision and representation : The controversial case of DSM-5. *Public Affairs Quarterly*, 29(1), 76-108.
- Smith, T. (2004). "Social" objectivity and the objectivity of values. Dans P. Machamer et G. Wolters (dir.), *Science, values, and objectivity* (p. 143-171). Pittsburgh : University of Pittsburgh Press.
- Sobel, M. et Hansen, K. (2006). Banking of human biological materials for research. Dans E. Bankert et R. Amdur (dir.), *Institutional review board : Management and function* (p. 433-440). Sudbury, MA : John and Bartlett.
- Solomon, M. (2006). Norms of epistemic diversity. *Episteme*, 3(1), 23-36.
- Solomon, M. (2007). *Social empiricism*. Cambridge, MA : MIT Press.
- Sommerhoff, G. (1950). *Analytical biology*. Oxford : Oxford University Press.
- Spitzer, R. L. et Endicott, J. (1978). Medical and mental disorder : Proposed definition and criteria. Dans R. L. Spitzer et D. F. Klein (dir.), *Critical issues in psychiatric diagnosis* (p. 15-39). New York : Raven Press.
- Spitzer, R. L. et Wakefield, J. C. (1999). DSM-IV diagnostic criterion for clinical significance : Does it help solve the false positives problem? *American Journal of Psychiatry*, 156(12), 1856-1864.
- Spitzer, R. L. (1981). The diagnostic status of homosexuality in DSM-III : A reformulation of the issues. *The American Journal of Psychiatry*, 138(2), 210-215.
- Spitzer, R. L. (1999). Harmful dysfunction and the DSM definition of mental disorder. *Journal of Abnormal Psychology*, 108(3), 430-432.

- Spitzer, R. L. (2009). DSM-V transparency : Fact or rhetoric? *Psychiatric Times*, 26(3).
- Stanford, K. (2017). Underdetermination of scientific theory. Dans E. N. Zalta (dir.), *The Stanford encyclopedia of philosophy*.
<https://plato.stanford.edu/archives/spr2017/entries/scientific-underdetermination/>
- Steel, D., Fazelpour, S., Gillette, K., Crewe, B. et Burgess, M. (2018). Multiple diversity concepts and their ethical-epistemic implications. *European Journal for Philosophy of Science*, 8(3), 761-780.
- Stein, D. J. (2014). Psychopharmacology and natural kinds : A conceptual framework. Dans S. Kincaid et K. Sullivan (dir.), *Classifying psychopathology : Mental kinds and natural kinds* (p. 65-73). Cambridge, MA : MIT Press.
- Stein, D. J. et Phillips, K. A. (2013). Patient advocacy and DSM-5. *BMC Medicine*, 11(1), 133-138.
- Szasz, T. S. (1960). The myth of mental illness. *American psychologist*, 15(2), 113-118.
- Szasz, T. (1961), *The myth of mental illness : Foundations of a theory of personal conduct*. New York : Harper & Row.
- Tabb, K. (2015). Psychiatric progress and the assumption of diagnostic discrimination. *Philosophy of Science*, 82, 1047-1058.
- Tekin, Ş. (2019). The missing self in scientific psychiatry. *Synthese*, 196, 2197-2215.
- Telles-Correia, D. (2018). Mental disorder : Are we moving away from distress and disability? *Journal of Evaluation in Clinical Practice*, 24(5), 973-977.

- Telles-Correia, D., Saraiva, S. et Gonçalves, J. (2018). Mental disorder : The need for an accurate definition. *Frontiers in Psychiatry*, 9(64). Récupéré de <https://www.frontiersin.org/articles/10.3389/fpsyt.2018.00064/full>.
- Thomas, A. et Sillen, S. (1972). *Racism and psychiatry*. New York : Carol Publishing Group.
- Thornton, T. (2007). *Essential philosophy of psychiatry*. Oxford : Oxford University Press.
- Tseris, E. (2018). Biomedicine, neoliberalism and pharmaceuticalisation of society. Dans A. V. Horwitz et A. Scull, *The Routledge international handbook of critical mental health* (p. 169-176). New York : Routledge.
- Tsou, J. Y. (2011). The importance of history for philosophy of psychiatry : The case of the DSM and psychiatric classification. *Journal of the Philosophy of History*, 5, 446-470.
- Tsou, J. Y. (2015). DSM-5 and psychiatry's second revolution : Descriptive vs. theoretical approaches to psychiatric classification. Dans S. Demazeux e P. Singy (dir.). *The DSM-5 in perspective* (p. 43-62). Dordrecht : Springer.
- Turnbull, M. G. (2018). Underdetermination in science : What it is and why we should care. *Philosophy Compass*, 13(2). doi.org/10.1111/phc3.12475
- Tyrer, P. et Steinberg, D. (2006). *Models for mental disorder : Conceptual models in psychiatry*. (4th edition). West Sussex : John Wiley & Sons.
- Ussher, J. M. (1991). *Women's madness : Misogyny or mental illness?* Londres : Harvester Wheatsheaf.
- Varga, S. (2011). Defining mental disorder : Exploring the "natural function" approach. *Philosophy, Ethics, and Humanities in Medicine*, 6(1), 1-10.
- Varga, S. (2015). *Naturalism, interpretation, and mental disorder*. New York : Oxford University Press.

- Varga, S. (2017). Mental disorder between naturalism and normativism. *Philosophy Compass*, 12(6), 1-9.
- Varga, S. (2018). Epistemic authority, philosophical explication, and the bio-statistical theory of disease. *Erkenntnis*, 85, 937-956.
- Wachbroit, R. (1994). Normality as a biological concept. *Philosophy of Science*, 61, 579-591.
- Wakefield, J. C. (1992a). Disorder as harmful dysfunction : A conceptual critique of DSM-III-R's definition of mental disorder. *Psychological Review*, 99(2), 232-247.
- Wakefield, J. C. (1992b). The concept of mental disorder : On the boundary between biological facts and social values. *American Psychologist*, 47(3), 373-388.
- Wakefield, J. C. (1997). Diagnosing DSM-IV-Part II : Eysenck (1986) and the essentialist fallacy. *Behaviour Research and Therapy*, 35(7), 651-665.
- Wakefield, J. C. (1999a). Mental disorder as a black box essentialist concept. *Journal of Abnormal Psychology*, 108(3), 465-472.
- Wakefield, J. C. (1999b). Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology*, 108(3), 374-399.
- Wakefield, J. C. (2001). Evolutionary history versus current causal role in the definition of disorder : Reply to McNally. *Behaviour Research and Therapy*, 39(3), 347-366.
- Wakefield, J. C. (2003). Dysfunction as a factual component of disorder : Reply to Houts, Part 2. *Behaviour Research and Therapy*, 41(8), 969-990.
- Wakefield, J. C. (2004). The myth of open concepts : Meehl's analysis of construct meaning versus black box essentialism. *Applied and Preventive Psychology*, 11(1), 77-82.

- Wakefield, J. C. (2005). On winking at the facts, and losing one's Hare : Value pluralism and the harmful dysfunction analysis. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, 4(2), 88-89.
- Wakefield, J. C. (2007). The concept of mental disorder : Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, 6(3), 149-156.
- Wakefield, J. C. (2013). Addiction, the concept of disorder, and pathways to harm : Comment on Levy. *Frontiers in Psychiatry*, 4(34).
doi.org/10.3389/fpsy.2013.00034
- Wakefield, J. C. (2015). Biological function and dysfunction : Conceptual foundations of evolutionary psychopathology. Dans David M. Buss (dir.), *The Handbook of Evolutionary Psychology*.
doi:10.1002/9781119125563.evpsych242
- Wakefield, J. C. et First, M. B. (2003). Clarifying the distinction between disorder and non-disorder : Confronting the overdiagnosis ("false positives") problem in DSM-V. Dans K. A. Phillips, M. B. First et H. A. Pincus (dir.), *Advancing DSM : Dilemmas in psychiatric diagnosis* (p. 23-56). Arlington, VA : American Psychiatric Publisher.
- Wakefield, J. C. et Conrad, J. A. (2019). Does the harm component of the harmful dysfunction analysis need rethinking? Reply to Powell and Scarffe. *Journal of Medical Ethics*, 45(9), 594-596.
- Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 104(12), 639-659.
- Whitbeck, C. (1981). A theory of health. Dans A. L. Caplan, H. T. Engelhardt et J. J. McCartney (dir.), *Concepts of health and disease : Interdisciplinary perspectives* (p. 611-626). Reading, VA : Addison-Wesley.
- Whooley, O. et Horwitz, A. V. (2013). The paradox of professional success : Grand ambition, furious resistance, and the derailment of the DSM-5 revision process. Dans J. Paris et J. Phillips (dir.), *Making the DSM-5* (p. 75-92). New York : Springer.

- Williams, B. (1985). *Ethics and the limits of philosophy*. Cambridge, MA : Harvard University Press.
- Williams, S. J., Martin, P. et Gabe, J. (2011). The pharmaceuticalisation of society? A framework for analysis. *Sociology of Health and Illness*, 33(5), 710-725.
- Wilson, M. (1993). DSM-III and the transformation of American psychiatry : A history ». *American Journal of Psychiatry*, 150, 399-410.
- Wimsatt, W. C. (1972). Teleology and the logical structure of function statements. *Studies in the History and Philosophy of Science*, 3, 1-80.
- Wouters, A. G. (2003). Four notions of biological function. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 34(4), 633-668.
- Wylie, A. (2003). Why standpoint matters. Dans R. Figueroa et S. Harding (dir.), *Science and other cultures : Issues in philosophies of science and technology*. New York : Routledge.
- Wylie, A. (2012). Feminist philosophy of science : Standpoint matters. *Proceedings and Addresses of the American Philosophical Association*, 86(2), 47-76.
- Wylie, A. (2015). A plurality of pluralisms : Collaborative practice in archaeology. Dans F. Padovani, A. Richardson et J. Y. Tsou (dir.), *Objectivity in science : New perspectives from science and technology studies*. (p. 189-210). Dordrecht : Springer.
- Yager, J. et McIntyre, J. S. (2014). DSM-5 clinical and public health committee : Challenges and considerations. *American Journal of Psychiatry*, 171(2), 142-144.